2001–2002 Technical Manual

Measured Progress

171 WATSON RD., DOVER, NH 03820 (800) 431-8901 WWW.MEASUREDPROGRESS.ORG

TABLE OF CONTENTS

CHAPTER 1—BACKGROUND AND OVERVIEW

PURPOSE OF THIS MANUAL

The purpose of this technical manual is to document the technical aspects of the 2001–2002 Maine Educational Assessment (MEA). In the fall of 2001, students in grades 4, 8, and 11 participated in the administration of the revised MEA in writing, reading, and health education. In the spring of 2002, students in grades 4, 8, and 11 were administered tests in mathematics, science and technology, social studies, and visual and performing arts. This report provides information about the technical quality of those assessments, including a description of the processes used to develop, administer, and score the tests and to analyze the test results. This report is intended to serve as a guide for replicating and/or improving the procedures in subsequent years.

While some parts of this technical report may be used by educated laypersons, the intended audience is experts in psychometrics and educational research. The report assumes a working knowledge of measurement concepts such as "reliability" and "validity," and statistical concepts such as "correlation" and "central tendency." In some chapters, the reader is presumed also to have basic familiarity with advanced topics in measurement and statistics.

LEARNING RESULTS

Following enactment of the Education Reform Act of 1984, Maine schools undertook a wide variety of initiatives designed to improve the quality of teaching and learning. Many of the lessons learned from those initiatives informed *Maine's Common Core of Learning*, a document published in 1990 that articulates a common vision for education in Maine by defining the knowledge, skills, and attitudes that all students should possess upon graduation from high school. In 1993, the Legislature directed the State Board of Education to undertake the next step in education reform by establishing a Task Force on *Learning Results* that was directed to

> "develop long-range education goals and standards for school performance and student performance to improve learning results and recommend to the commissioner and to the Legislature a plan for achieving those goals and standards."

After substantial work, in January of 1996 the Task Force presented a report to the Legislature that contained a series of recommendations together with a set of standards, a plan for implementation, and proposed legislation. After a series of intense hearings during the 1996 Legislative Session, the Legislature adopted much of the work of the Task Force and directed the Department of Education and the State Board of Education to continue to develop the *Learning Results.*

Acting on the recommendations of the Task Force, the Legislature adopted six Guiding Principles that describe the characteristics of a well-educated person. To fulfill these principles, the Legislature required that the Department of Education and the State Board of Education develop *Learning Results* within the following eight areas:

Career Preparation
English Language Arts
Health and Physical Education
Mathematics
Modern and Classical Languages
Science and Technology
Social Studies
Visual and Performing Arts

These are not "subjects" in the same sense that the word is used when referring to courses in school. They are areas of learning that will in some cases cut across a number of discrete courses or disciplines. In response to the legislative directive, the Commissioner appointed a working group, known as the Critical Review Committee, to prepare a draft of standards for consideration by the State Board of Education and by the Legislature. The Committee met on numerous occasions during the summer and fall of 1996 to produce this revised document, which was approved in May of 1997 by the 118th Legislature.

PURPOSES OF THE MEA

The *Learning Results* are just one part of an educational system. As goals for what all students should know and be able to do upon finishing school, they are not written to prescribe a minimum of "passing" standard. The

setting of minimum requirements is the function of assessments that are separate from the creation of academic goals.

Because some students are ready for assessment at earlier stages than others, no assumption is made about when a standard might be achieved.

"The statute passed in April of 1996 includes the following provisions relating to assessment:

Student achievement of the learning results. . .must be measured by a combination of state and local assessments to measure progress and ensure accountability. The 4th-grade, 8th-grade, and 11th-grade results of the Maine Educational Assessment, the "MEA," are the state assessments used to measure achievement of the learning results. The 4th-grade and 8th-grade MEA must be used to measure achievement of the learning results beginning in the 1998-99 school year. Local school administrative units may develop additional assessments to measure achievement of the learning results, including student portfolios, performances, demonstrations, and other records of achievements."

An Assessment Design Team comprised of Maine educators and assessment specialists has been established to redesign state level assessments and to assist in development of high-quality local assessments that will be used to measure student achievement of the *Learning Results*. The statewide assessment system they are developing will

- align with Maine's *Learning Results;*
- utilize multiple measures of learning;
- ensure fair and equitable assessment for all students;
- utilize recognized, relevant technical standards for assessment;
- provide understandable information to educators, parents, students, the public, and the media;
- provide professional development opportunities for teachers, administrators, and future educators; and
- be practical and manageable.

ORGANIZATION OF THIS MANUAL

The organization of this manual is based on the conceptual flow of an assessment's life span; it begins with the initial test specification and addresses all the intermediate steps that lead to final score reporting. Section I covers

the development of the MEA tests. It consists of eight chapters covering general design issues; the test development process; and the specific designs of the English language arts, mathematics, science and technology, social studies, visual and performing arts, and health education assessments. Section II consists of a single chapter describing the administration of the tests. Section III contains six chapters covering scoring, equating and scaling, item analysis, reliability, validity, and reporting. Section IV contains references and Section V contains the appendices.

SECTION I: ASSESSMENT DEVELOPMENT

CHAPTER 2—OVERVIEW OF TEST DESIGN

LEARNING RESULTS

MEA questions are directly linked to the **content standards** and **performance indicators** described in Maine's *Learning Results*. The content standards are the basis for the reporting categories developed for each subject area; the performance indicators are used to help guide the development of test questions. No other content or process is subject to statewide assessment. An item may address part, several, or all of the performance indicators.

ITEM TYPES

Maine's educators and students were familiar with the item types that were used in the 2001-02 assessment program as all had been previously introduced. The item types used and the functions of each are described below.

Multiple-choice items were used, in part, to provide breadth of coverage of a subject area. Because they require no more than a minute for most students to answer, these items make efficient use of limited testing time and allow coverage of a wide range of knowledge and skills.

Short-answer items were used to assess students' skills and their abilities to work with brief, well-structured problems that had one or a very limited number of solutions (e.g., mathematical computations). Short-answer items require approximately two to five minutes for most students to answer. The advantage of this item type is that it requires students to demonstrate knowledge and skills by generating, rather than merely selecting, an answer.

Constructed-response items typically require students to use higher-order thinking skills—evaluation, analysis, summarization, and so on—in constructing a satisfactory response. Constructed-response items should take most students approximately five to ten minutes to complete. It should be noted that the use of previously released MEA items prepare students to answer this kind of item was appropriate and encouraged.

Extended-response items were discontinued in the 2001-02 MEA and each one was replaced with two constructed-response items. This change resulted in the same number of score points and provided improved distribution of test items across the content standards.

COMMON-MATRIX DESIGN

The 2001-02 MEA continued to measure what students know and are able to do by using a variety of test item types. The tests continued to be structured using both *common* and *matrix-sampled* items. Common items are those taken by all students at a given grade level; in addition, a larger pool of matrix-sampled items is divided among the multiple forms of the test at each grade level. Each student took only one form of the test and so answered a fraction of the matrix-sampled items in the entire pool. This design, which has been used throughout the MEA's history, provides reliable and valid results at the student level. It also provides greater breadth of coverage of a content area for school results while minimizing testing time.

In 2001–02, the reports continued to report out only common scores in the student level results for ease of understanding them. If student results were based on common and matrix-sampled items, one student could score higher than another in raw score, but lower in scaled score. By producing common results only, this type of reversal was avoided.

EMBEDDED FIELD TEST

Historically, MEA field tests were administered independently of the statewide assessment at schools willing to participate. In recent years, however, it has become increasingly difficult to find schools that have the time to field test, and sample sizes were decreasing to an undesirable level. The 2001-02 MEA was designed to include an embedded field test in all content areas that was transparent to test takers and that had a negligible impact on testing time. Because the field test was taken by all students, it provided the sample needed to produce reliable data with which to inform item selection for future tests.

The embedded field test achieved two other objectives. First, it created a pool of replacement items needed due to natural attrition caused by the release of all common items each year in English language arts, science and technology, social studies, and mathematics. Second, the embedded field test ensured that there would be sufficient

numbers of items to fill the gaps in coverage of the standards and performance indicators that result when common items are released and matrix items move to common. While the health education and visual and performing arts assessments are matrix-sampled only, three constructed-response health items and two constructed-response visual and performing arts items were also released from the 2001-02 MEA.

TEST BOOKLET DESIGN

In order to accommodate the embedded field test for the fall English language arts, writing, and health assessments, there were 16 unique test forms at each grade. Forms 1 through 10 contained the common and matrix portions of the test, and forms 11 through 16 were sub-forms that contained the common and embedded field test items in place of the matrix items. This design allowed administration of the field test without lengthening testing time and was necessary due to the unique structure of the English language arts test that is dependent upon reading passages. While it is true that not every student took the field test, the sample size was approximately 500 students and thus yielded sufficient data with which to make item selections.

The spring administration for the science and technology, social studies, mathematics, and visual and performing arts assessments comprised the usual number of 12 unique forms. In this administration, every student took the embedded field test. However, only the responses of the students in the same schools that took the fall embedded field test were scored.

For the 2001-02 MEA, it was decided that the grade 4 integrated test and response booklet should be separated into two documents for the spring administration only. Publishing a separate 32-page scannable response document instead of an 80-page combined scannable test and response document resulted in substantial savings to the DOE.

TEST SESSION TIMES

The MEA tests were given at two different times during the school year: **writing, reading, and health education** were administered to all grades in late fall, and tests in **mathematics, science and technology, social studies, and visual and performing arts** were administered to all grades during a two-week period in early March. Schools were able to schedule testing sessions at any time during the first week of this period, provided they followed the sequence in the scheduling guidelines detailed in test administration manuals and that all testing classes

within a school were on the same schedule. The second week was reserved for make-up testing of students who were absent from initial test sessions.

The timing and scheduling guidelines for MEA tests were based on estimates of the time it would take an average student to respond to each type of item that makes up the test:

- multiple-choice– 1 minute
- short-answer– 2 minutes
- constructed-response– 10 minutes

For the English language arts reading test, the scheduling guidelines included an estimate of 10 minutes to read each passage used in the assessment.

While the guidelines for scheduling are based on the assumption that most students will complete the test within the time estimated, each test session was scheduled so that additional time was provided for students who needed it. One-third additional time was allocated for each session (i.e., 45-minute sessions with an additional 15 minutes and 35-minute sessions with an additional 10 minutes).

If additional classroom space was not available for students who required additional time to complete the tests, schools were allowed to consider using another space, such as the guidance office, for this purpose. If additional areas were not available, it was recommended that each classroom being used for test administration be scheduled for the maximum amount of time. Detailed instructions on test administration and scheduling were provided in the coordinator's and administrator's manuals.

CHAPTER 3—TEST DEVELOPMENT PROCESS

DEVELOPMENT COMMITTEE ITEM IDEA GENERATION

The development of the MEA tests continues to be a cooperative effort by content development committees comprising Maine teachers, curriculum supervisors, higher education faculty, content specialists of the Department of Education, and curriculum/assessment specialists employed by the program's contractor, Measured Progress. The committees are structured to represent all areas of the state and committee members all serve rotating terms.

The committees' primary roles are to develop test items for the MEA and to interpret testing data so that those items can be selected for the program. The 2001-02 MEA development committee for each subject area at grade levels 4, 8, and 11 met two times. In the first meeting, after reviewing the content standards and test specifications, committee members approved which items from the 2000-01 MEA would move to common. Then they brainstormed or drafted test items and scoring rubrics for the embedded field test items that would fill the gaps in coverage of the standards left after items moved to common. In the second meeting, the committees reviewed item statistics and made recommendations about selecting, revising, or eliminating specific items from the item pool for the operational test. At that time, the committees also confirmed that each item aligned directly to Maine's *Learning Results* and was assigned to the appropriate content standard reported in school and district results. Because all common MEA items are released to the public each year, the committees repeat these activities annually as new items are developed in order to replenish the item pool.

INTERNAL ITEM REVIEW

- The lead or peer test developer within the content specialty reviewed the typed item, constructed-response scoring guide, and any reading selections and graphics.
- The content reviewer considered item "integrity;" item content and structure; appropriateness to designated content area; item format; clarity; possible ambiguity; keyability; single "keyness;" appropriateness and quality of reading selections and graphics; and appropriateness of scoring guide descriptions and distinctions (as correlated to the item and within the guide itself).
- The content reviewer also considered scorability and evaluated whether the scoring guide adequately addressed performance on the item.

- Fundamental questions the content reviewer considered, but was not limited to, included the following:
  - What is the item asking?
  - Is the key the only possible key? (Is there only *one* correct answer?)
  - Is the constructed-response item scorable as written (were the correct words used to elicit the response defined by the guide)?
  - Is the wording of the scoring guide appropriate and parallel to the item wording?
  - Is the item complete (e.g., with scoring guide, content codes, key, grade level, and contract identified)?
  - Is the item appropriate for the designated grade level?

EXTERNAL ITEM REVIEW

- Item sets were brought to Content Development Committee meetings for review and revision.

ITEM EDITING

Editors reviewed and edited the items from the Content Development Committee item review to ensure uniform style (based on *The Chicago Manual of Style, 14th Edition*) and adherence to sound testing principals. These principles included the stipulation that items

- were correct with regard to grammar, punctuation, usage, and spelling;
- were written in a clear, concise style;
- contained unambiguous explanations to students as to what is required to attain a maximum score;
- were written at a reading level that would allow the student to demonstrate his or her knowledge of the tested subject matter, regardless of reading ability;
- exhibited high technical quality regarding psychometric characteristics;
- had appropriate answer options or score-point descriptors; and
- were free of potentially sensitive content.

REVIEWING AND REFINING

Test developers presented item statistics to the development committees to assist in the committees' recommendations for placement of items into the common and matrix portions of the test. The Department of Education made the final selections with the assistance of Measured Progress at a meeting.

OPERATIONAL TEST ASSEMBLY

Test assembly is the sorting and laying out of item sets into test forms. Criteria considered during this process included the following:

- **Content coverage/match to test design.** The curriculum specialist completed an initial sorting of items into sets based on a balance of content categories across sessions and forms, as well as a match to the test design (e.g., number of multiple-choice, short-answer, constructed-response, and extended-response items).
- **Item difficulty and complexity.** Item statistics drawn from the data analysis of previously tested items were used to ensure that there were similar levels of difficulty and complexity across forms.
- **Visual balance.** Item sets were reviewed to ensure that each reflected a similar length and "density" of selected items (e.g., length/complexity of reading selections, or number of graphics).
- **Option balance.** Each item set was checked to verify that it contained a roughly equivalent number of key options (As, Bs, Cs, and Ds).
- **Name balance.** Item sets were reviewed to ensure that a diversity of names was used.
- **Bias.** Each item set was reviewed to ensure fairness and balance based on gender, ethnicity, religion, socio-economic status, and other factors.
- **Page fit.** Item placement was modified to ensure the best fit and arrangement of items on any given page.
- **Facing page issues.** For multiple items associated with a single stimulus (a graphic or reading selection), consideration was given to whether those items needed to begin on a left- or right-hand page, as well as to the nature and amount of material that needed to be placed on facing pages. These considerations served to minimize the amount of "page flipping" required of the students.
- **Relationships between forms.** Sets of common items were placed identically in each version of the forms. Although matrix-sampled item sets differ from form to form, they must take up the same number of pages in

each form so that sessions and content areas begin on the same page in every form. Therefore, the number of pages needed for the longest form often determines the layout of each form.

- **Visual appeal.** The visual accessibility of each page of the form was always taken into consideration, including such aspects as the amount of "white space," the density of the text, and the number of graphics.

EDITING DRAFTS OF OPERATIONAL TESTS

Any changes made by the test construction specialist must be reviewed and approved by the test developer. Once a form had been laid out in what was considered its final form, it was reread to identify any final considerations, including the following:

- Editorial changes. All text was scrutinized for editorial accuracy, including consistency of instructional language, grammar, spelling, punctuation, and layout. Measured Progress' publishing standards are based on *The Chicago Manual of Style, 14th Edition*.
- "Keying" items. Items were reviewed for any information that might "key" or provide information that would help answer another item. Decisions about moving keying items are based on the severity of the "key-in" and the placement of the items in relation to each other within the form.
- Key patterns. The final sequence of keys was reviewed to ensure that their order appeared random (e.g., no recognizable pattern, and no more than three of the same key in a row).

BRAILLE AND LARGE-PRINT TRANSLATION

Form 1 for the grades 4, 8, and 11 tests was translated into Braille by a subcontractor that specializes in test materials for blind and visually impaired students. In addition, Form 1 for each grade was adapted into a large-print version.

CHAPTER 4—DESIGN OF ENGLISH LANGUAGE ARTS ASSESSMENT

READING

BLUEPRINT

As indicated earlier, the English language arts framework for reading is based on Maine's *Learning Results*, which identifies five **content standards** that apply specifically to reading and reading comprehension. Those content standards are:

- **Process of reading:** Students use the skills and strategies of the reading process to comprehend, interpret, evaluate, and appreciate what they have read.
- **Literature and culture:** Students use reading, listening, and viewing strategies to experience, understand, and appreciate literature and culture.
- **Language and images:** Students demonstrate an understanding of how words and images communicate.
- **Informational texts:** Students apply reading, listening, and viewing strategies to informational texts across all areas of curriculum.

The content standards have been adapted to create a reporting category framework for reading, as shown below.

Passage Type	Comprehension of Literary and Informational Texts			Total
	Reading Comprehension and Literary Analysis	A. Process of Reading	C. Language and Images	
B. Literature and Culture: Literary Passages				50%
D. Informational Texts: Content Passages Practical Passages				50% (30%) (20%)
Total	80%	20%		100%

CONTENT SPECIFICATIONS

The first major reporting category at the student, school, and district levels is "comprehension of literary and informational texts." The data generated for this reporting category were based on items related to three types of reading passages that reflect standards B and D of the English Language Arts (ELA) *Learning Results*. The passage

types were identical to those that have been used in the MEA in past years. Fifty percent of the passages comprised literary works; 30% were selected from content pieces (see explanation below); and 20% were drawn from practical sources (see explanation below).

Passages included both long and short "authentic" texts selected from reading sources that students at each grade level would be likely to encounter in their classroom and in their independent reading. The passages were not written specifically for the assessment, but instead were collected from published works.

- **Literary passages** are represented by a variety of genres—modern narratives; diary entries; drama; poetry; biographies; essays; excerpts from novels; short stories; and traditional narratives, such as fables, myths, and folktales.
- **Content passages** are primarily informational and often deal with the areas of science and social studies. They are drawn from such sources as newspapers, magazines, and books.
- **Practical passages** are functional materials that instruct or advise the reader—for example, directions, reference tools, or manuals.

The main difference in the passages used for grades 4, 8, and 11 is the degree of difficulty. All passages were selected to be appropriate for the intended audience; however, the ideas expressed become increasingly more complex at grade levels 8 and 11.

The items related to these passages require students to demonstrate their skills in both literal comprehension (where the answer is stated explicitly in the text) and inferential comprehension (where the answer is implied by the text and/or the text must be connected to relevant prior knowledge to determine an answer). In addition, some items focus on the reading skills reflected in content standards A and C of the *Learning Results*. Items of this type require students to use the skills and strategies of reading to answer items—for example, how to identify the author's principal purpose, such as to persuade, entertain, or inform—and to demonstrate their understanding of how words and images communicate to readers.

ITEM TYPES

The MEA English language arts assessment in reading included multiple-choice, short-answer, and constructed-response items, as well as one extended-response/writing sample item. Short-answer items required

students to write an answer consisting of several phrases or short sentences. Each type of item was worth a specific number of points in the student's total language arts score, as shown below.

Type of Item	Possible Score Points
Multiple Choice	0–1
Short Answer	0–2
Constructed Response	0–4
Extended Response/Writing Sample	0–4

TEST DESIGN

The table below summarizes the numbers and types of items that were used in the MEA reading assessment for 2001-02.

Session	COMMON				MATRIX				Time (minutes)
	MC	SA	CR	ER	MC	SA	CR	ER	
2A	8	1	2	0					35 (+10)
2B	8	2	2	0					35 (+10)
3A	8	1	0	1					35 (+10)
3B					8	0	2	0	35 (+10)

Key

- MC = multiple-choice
- SA = short-answer
- CR = constructed-response
- ER = extended-response/writing sample

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

ENGLISH LANGUAGE ARTS — READING
GRADE 4

	Standard A				Standard B				Standard C				Standard D				Total
Common Passages	**MC** x 1	**SA** x 2	**CR** x 4	**Pts**	**MC** x 1	**SA** x 2	**CR** x 4	**Pts**	**MC** x 1	**SA** x 2	**CR** x 4	**Pts**	**MC** x 1	**SA** x 2	**CR** x 4	**Pts**	**Points 212**
Cleaning Up the Ocean		1		2	2		1	6				0	2			2	10
Drinking Milk Is Good for Birds	2		1	6	1			1	1		1	5	4	1		6	18
I'm Going To Be Famous				0	7	1	1	13	1			1				0	14
One Brave Summer				0	2	1	1	8	2			2				0	10
Matrix Passages																	
Alcove Spring	2			2	4		2	12	2			2				0	16
Amazing Spiders	1			1				0				0	3		1	7	8
Avalanche!	2			2	6		2	14				0				0	16
Bacon-Tomato Sandwiches	1			1				0				0	3		1	7	8
Be a Junk Food Detective	2			2				0				0	2		1	6	8
Brian's Winter	2			2	1		1	5	1			1				0	8
Climbing/Every Time I Climb a Tree				0	2		1	6	2			2				0	8
Let's Write a True Life Story	1			1				0	1			1	2		1	6	8
Marsha	1			1	1		1	5	2			2				0	8
My Dino Discovery	1			1				0	1			1	2		1	6	8
Ruby	1			1	2		1	6	1			1				0	8
Secret Place				0	3		1	7	1			1				0	8
Should Your School Tell You What To Wear	3			3	1		2	9				0	4			4	16
The Death of the Princess's Cat	1			1	7		2	15				0				0	16
Welcome to the Inventor's Club	2			2				0	1			1	5		2	13	16

ENGLISH LANGUAGE ARTS — READING
GRADE 8

	Standard A				Standard B				Standard C				Standard D				Total Points 212
Common Passages	**MC** x 1	**SA** x 2	**CR** x 4	**Pts**	**MC** x 1	**SA** x 2	**CR** x 4	**Pts**	**MC** x 1	**SA** x 2	**CR** x 4	**Pts**	**MC** x 1	**SA** x 2	**CR** x 4	**Pts**	
Go Fly a Kite	1		1	5				0	2			2	5	1	1	11	18
Graduation Morning				0	3		1	7	1	1		3				0	10
Niagara Falls	1			1	3			3				0		1	1	6	10
Uncle Joe	2			2	6	1	1	12				0				0	14
Matrix Passages																	
Amir	1			1	7		2	15				0				0	16
Cool Science-A Lesson Runs Through It			1	4				0	1			1	3			3	8
Diary of Anne Frank/Zlata's Diary				0	2		1	6	2			2				0	8
First Lesson/Fathers				0	4		1	8				0				0	8
Gentle Friends, Essential Allies				0				0				0	4		1	8	8
Hurricanes				0				0	2			2	6		2	14	16
Lost in the Woods	2			2				0				0	6		2	14	16
Right Smart O' Wind	1		1	5	3		1	7	4			4				0	16
Road Runner Sports	2			2				0				0	2		1	6	8
Springsteen Concert Debated				0	1			1	2			2	5		2	13	16
The Life of the Ladybird Beetle	1		2	9	1			1	2			2	4			4	16
Why I Never Shoot Bears				0	3		1	7	1			1				0	8
Wreck of the Monkey Cage				0	3		1	7	1			1				0	8
You Can Be an Inventor	3			3				0	1			1			1	4	8

ENGLISH LANGUAGE ARTS — READING
GRADE 11

	Standard A				Standard B				Standard C				Standard D				Total Points 212
Common Passages	MC x 1	SA x 2	CR x 4	Pts	MC x 1	SA x 2	CR x 4	Pts	MC x 1	SA x 2	CR x 4	Pts	MC x 1	SA x 2	CR x 4	Pts	
A Presidential Candidate	2			2	5	1	1	11	1			1				0	14
Chief Joseph of the Nez Perce	2			2				0				0	2	1	1	8	10
Reading Moving Water	2			2				0	1			1	5	1	2	15	18
The House on Mango Street				0	2	1	1	8	2			2				0	10
Matrix Passages																	
A Day at the Theater	1			1				0	2			2	5		2	13	16
At Harvesttime				0	3			3	1		1	5				0	8
Children of the Sun	1			1	3		1	7				0				0	8
Discover Whitewater Rafting	1			1				0				0	3		1	7	8
Feet	1			1				0	1			1	2		1	6	8
I Wandered Lonely as a Cloud				0	2		1	6	2			2				0	8
Life in the Thirteen Colonies	1			1				0				0	3		1	7	8
Nearer	1			1	2		1	6	1			1				0	8
Piltdown Man	1			1				0				0	3		1	7	8
Polonius's Advice to Laertes	2			2	2		1	6				0				0	8
Prevent Repetitive Strain at the Keyboard	2			2				0				0	2		1	6	8
Snails and Slugs	2			2				0	1			1	5		2	13	16
The Country of the Pointed Firs	2			2	5		2	13	1			1				0	16
The Ojibwa Corn Hero	1		1	5	7		1	11				0				0	16
Why You Like Some Food and Hate Others	4			4				0				0	4		2	12	16

WRITING

BLUEPRINT

The MEA assesses students' writing skills directly through the use of writing prompts, or topics, to which students respond. Maine's *Learning Results* includes two content standards that apply specifically to writing. Those content standards are

- **Standard English conventions:** Students write and speak correctly, using conventions of standard written and spoken English.
- **Stylistic and rhetorical aspects of writing and speaking:** Students use stylistic and rhetorical aspects of writing and speaking to explore ideas, to present lines of thought, to represent and reflect on human experience, and to communicate feelings, knowledge, and opinions.

The *Learning Results* standards were adapted to create reporting categories for writing, as shown below.

Stylistic and Rhetorical Aspects of Writing	▪ Idea/topic development ▪ Organization ▪ Supporting detail
Standard English Conventions	▪ Grammar ▪ Spelling ▪ Punctuation ▪ Capitalization ▪ Sentence structure

CONTENT SPECIFICATIONS

Four broad types, or modes, of writing are used in the MEA, as listed below[1]:

- **Narration:** Narrative writing answers the question, "What happened?" It tells a story through a sequence of events, so that the reader understands the action.
- **Exposition:** Expository writing informs the reader about something. Methods of exposition include comparison and contrast, illustration, classification, definition, and analysis. Methods of exposition are often combined to accomplish a specific purpose for writing.
- **Description:** Descriptive writing presents the qualities of objects, persons, conditions, and actions.

[1] Descriptions are adapted from *Modern Rhetoric*, by Cleanth Brooks and Robert Penn Warren.

- **Persuasion/argument:** Persuasive writing uses emotional appeals to bring about a change of attitude, point of view, or feeling. Argumentative writing uses logic and reason to bring about a change of attitude, point of view, or feeling; it shows that a conclusion merits belief because of credible data, evidence, and so on.

The student's "audience" and "purpose for writing" also influence the development, style, and tone of a written composition. These were specified as part of the prompts and varied by grade level. In addition, the prompts were developed with the following criteria as guidelines:

- the prompts must be interesting to students;
- the prompts must be accessible to all students (i.e., all students would have something to say about the topic); and
- the prompts must generate sufficient text to be effectively scored.

The prompts used in the 2000-01 MEA writing assessment follow.

Grade 4 prompt: Write about a time when you experienced very good or very bad weather.

Grade 8 prompt: Describe a good friend.

Grade 11 prompt: Respond to one of the following quotes:

"Hitch your wagon to a star."—Ralph Waldo Emerson

"I never let schooling interfere with my education." —Mark Twain

TEST DESIGN

Each student responded to one common writing prompt, as well as a common extended-response/writing sample question that was scored for both reading and writing. The chart that follows outlines the total number of possible points—as reported—by learning results and item type.

ENGLISH LANGUAGE ARTS—WRITING
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common Prompt	Extended Response Writing	Total Points
Standard English Conventions (Standard F)	8	4	12
Stylistic and Rhetorical Aspects of Writing (Standard G)	12	6	18

Number of Points Possible
Grade 8

Standard	Common Prompt	Extended Response Writing	Total Points
Standard English Conventions (Standard F)	8	4	12
Stylistic and Rhetorical Aspects of Writing (Standard G)	12	6	18

Number of Points Possible
Grade 11

Standard	Common Prompt	Extended Response Writing	Total Points
Standard English Conventions (Standard F)	8	4	12
Stylistic and Rhetorical Aspects of Writing (Standard G)	12	6	18

CHAPTER 5—DESIGN OF THE MATHEMATICS ASSESSMENT

BLUEPRINT

The mathematics framework was based on Maine's *Learning Results*, which identifies eleven **content standards** as shown below:

- **Numbers and number sense:** Students understand and demonstrate a sense of what numbers mean and how they are used.
- **Computation:** Students understand and demonstrate computation skills.
- **Data analysis and statistics:** Students understand and apply concepts of data analysis.
- **Probability:** Students understand and apply concepts of probability.
- **Geometry:** Students understand and apply concepts from geometry.
- **Measurement:** Students understand and demonstrate measurement skills.
- **Patterns, relations, and functions:** Students understand that mathematics is the science of patterns, relationships, and functions.
- **Algebra concepts:** Students understand and apply algebraic concepts.
- **Discrete mathematics:** Students understand and apply concepts in discrete mathematics.
- **Mathematical reasoning:** Students understand and apply concepts of mathematical reasoning.
- **Mathematical communication:** Students reflect upon and clarify their understanding of mathematical ideas and relationships.

These standards were used to create a reporting category framework for mathematics, shown below. The framework was divided into two major areas:

- **content**, which refers to the student's knowledge and conceptual and procedural understanding of each standard, and
- **application**, which refers to a student's use of knowledge and conceptual and procedural understanding as a basis for application through reasoning, inquiry, communication of ideas, and problem solving.

Each item in the mathematics assessment measured a content standard; in addition, each item was reported as measuring either content or application.

As shown in the table below, the goal for distribution of items, or emphasis, across standards varies from grade to grade.

Content Standard	Grade		
	4	8	11
A. Number and Number Sense	15%	14%	10%
B. Computation	15%	11%	5%
C. Data Analysis and Statistics	12%	11%	10%
D. Probability	8%	11%	10%
E. Geometry	12%	11%	15%
F. Measurement	12%	10%	10%
G. Patterns, Relations, Functions	12%	13%	15%
H. Algebra Concepts	9%	14%	15%
I. Discrete Mathematics	5%	5%	10%

CONTENT AND APPLICATION

For students to function effectively as mathematical problem-solvers, they must be taught how to apply and communicate basic concepts and procedures as well as how to do the procedures. **Content items** measure what students have been taught directly, including the basic concepts and procedural skills from all the content standards. For example, in the numbers and number sense standard and the computation standard, conceptual and procedural knowledge includes understanding of place value in our number system; the computational algorithms as applied to whole numbers, fractions, and decimals; and the concepts of ratio, proportion, and percent. In the data analysis and statistics standard, conceptual and procedural knowledge includes the reading of charts and graphs as well as the concepts of averages (means, medians, and modes) and methods for computing them. Contextual settings used in items measuring this category are very simple and are directly related to those used in the teaching of the concepts and procedures.

Application items measure what the students can do with what they have been taught. Included are items requiring students to combine the basic concepts and procedures to solve real-life and mathematical problems, to evaluate their own ideas and the ideas of others using mathematical reasoning, and to communicate their ideas using the wealth of symbolic, pictorial, graphic, and verbal representations available in mathematics.

It is important to understand that application items also measure mastery of the basic concepts and procedures. For example, in mathematics, 52 percent of the items are either short-answer or constructed-response items (see "Content Specifications" below), which are worth up to 2 and 4 score points respectively. In most cases, portions of these items require the student to perform some problem solving, reasoning, and/or communicating, and so the items are classified under applications. At the same time, however, the items require students to demonstrate their understanding of mathematics content. If a student does not show mastery of all aspects of a short-answer or constructed-response item, or if he/she makes careless errors, the student does not earn the highest score for that item. Thus, it can be said that **all** mathematics items in the MEA measure content; some items go beyond that realm, however, and are classified for reporting purposes as application.

CONTENT SPECIFICATIONS

The MEA mathematics assessment included multiple-choice, short-answer, and constructed-response items. Each item type was worth a specific number of points in the student's total mathematics score, as shown below.

Type of Item	Possible Score Points
Multiple Choice	0–1
Short Answer	0–2
Constructed Response	0–4

TEST DESIGN

The tables below summarize the numbers and types of items that were used in the MEA mathematics assessment for 2001-02. The tables show the construction of the common and matrix-sampled portions of the assessment.

GRADE 4

Session	COMMON			MATRIX			Time (minutes)
	MC	SA	CR	MC	SA	CR	
4A (NC)	6	5	0	2	1	1*	30 (+10)
4B (C)	5	0	3	0	0	0	30 (+10)
4C (C)	9	0	1	4	0	1*	30 (+10)

*alternating matrix and field test item

GRADES 8 AND 11

Session	COMMON			MATRIX			Time (minutes)
	MC	SA	CR	MC	SA	CR	
4A NC	5	5	2	2	1	1*	50 (+15)
4B (C)	15	0	2	4	0	1*	50 (+15)

*alternating matrix and field test item

Key

- (C) = calculator use allowed
- (NC) = no calculator use allowed
- MC = multiple-choice
- SA = short-answer
- CR = constructed-response

THE USE OF CALCULATORS IN THE MEA

The Maine educators who designed and developed the assessment test acknowledge the importance of mastering of arithmetic algorithms. At the same time, they understand that the use of calculators is a necessary and important skill in society today. Calculators can save time and error in the measurement of some higher order thinking skills and allow students to do more sophisticated and intricate problems. For these reasons, it was decided that calculators should be permitted in some parts of the MEA mathematics assessment and prohibited in others. (Students were allowed to use any calculator with which they are familiar.)

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

MATHEMATICS
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common				Matrix Per Form				Total Points 190
	MC x 1	SA x 2	CR x 4	Points	MC x 1	SA x 2	CR x 4	Points	
Content	15			15	64	3		70	85
Application	5	5	4	31	8	9	12	74	105
Numbers and Number Sense (Standard A)	3		1	7	14	2	1	22	29
Computation (Standard B)	5	1		7	8	3	2	22	29
Data Analysis and Statistics (Standard C)	4	1		6	9		2	17	23
Probability (Standard D)	1			1	8		1	12	13
Geometry (Standard E)	2		1	6	11		1	15	21
Measurement (Standard F)	2		1	6	7	3	1	17	23
Patterns, Relations, Functions (Standard G)	1		1	5	7	1	2	17	22
Algebra Concepts (Standard H)	2	1		4	5	2	1	13	17
Discrete Mathematics (Standard I)		2		4	3	1	1	9	13

MATHEMATICS
NUMBER OF POINTS POSSIBLE
GRADE 8

Standard	Common				Matrix Per Form				Total Points 109
	MC x 1	SA x 2	CR x 4	Points	MC x 1	SA x 2	CR x 4	Points	
Content	16	3	1	26	37	4	2	53	79
Application	4	2	3	20	35	8	10	91	111
Numbers and Number Sense (Standard A)	2	2		6	16		1	20	26
Computation (Standard B)	5			5	9	1	1	15	20
Data Analysis and Statistics (Standard C)	1		1	5	7	3	1	17	22
Probability (Standard D)	3	1		5	4	1	2	14	19
Geometry (Standard E)	1		1	5	4	2	2	16	21
Measurement (Standard F)	1		1	5	6	1	2	16	21
Patterns, Relations, Functions (Standard G)	2		1	6	11	3	1	21	27
Algebra Concepts (Standard H)	4	1		6	10	1	2	20	26
Discrete Mathematics (Standard I)	1	1		3	5			5	8

MATHEMATICS
NUMBER OF POINTS POSSIBLE
GRADE 11

Standard	Common				Matrix Per Form				Total Points 190
	MC x 1	SA x 2	CR x 4	Points	MC x 1	SA x 2	CR x 4	Points	
Content	8			8	28	3		34	42
Application	12	5	4	38	44	9	12	110	148
Numbers and Number Sense (Standard A)		1		2	4	1		6	8
Computation (Standard B)	6			6	13		1	17	23
Data Analysis and Statistics (Standard C)	1	1	1	7	10	2	1	18	25
Probability (Standard D)	2	1		4	7		2	15	19
Geometry (Standard E)	2		1	6	12	2	1	20	26
Measurement (Standard F)	3		1	7	5	1	1	11	18
Patterns, Relations, Functions (Standard G)	2		1	6	9	2	1	17	23
Algebra Concepts (Standard H)	3	1		5	9	3	4	31	36
Discrete Mathematics (Standard I)	1	1		3	3	1	1	9	12

CHAPTER 6—DESIGN OF THE SCIENCE AND TECHNOLOGY ASSESSMENT

BLUEPRINT

The science and technology framework was based on Maine's *Learning Results*, which identify thirteen **content standards** as listed below:

- **Classifying life forms:** Students understand that there are similarities within the diversity of all living things.
- **Ecology:** Students understand how living things depend on one another and on non-living aspects of the environment.
- **Cells:** Students understand that cells are the basic units of life.
- **Continuity and change:** Students understand the basis for all life and that all living things change over time.
- **Structure of matter:** Students understand the structure of matter and the changes it can undergo.
- **The Earth:** Students gain knowledge about the Earth and the processes that change it.
- **The universe:** Students gain knowledge about the universe and how humans have learned about it, and the principles upon which it operates.
- **Energy:** Students understand concepts of energy.
- **Motion:** Students understand the motion of objects and how forces can change that motion.
- **Inquiry and problem solving:** Students apply inquiry and problem-solving approaches in science and technology.
- **Scientific reasoning:** Students learn to formulate and justify ideas and to make informed decisions.
- **Communication:** Students communicate effectively in the applications of science and technology.
- **Implications of science and technology:** Students understand the historical, social, economic, environmental, and ethical implications of science and technology.

Nine of these standards (A through I) address the various content areas in science and technology as shown below.

Content Standard	Grade		
	4	8	11
A. Classifying Life Forms	10%	10%	8%
B. Ecology	12%	10%	10%
C. Cells	10%	15%	12%
D. Continuity and Change	10%	10%	12%
E. Structure of Matter	8%	15%	15%
F. The Earth	10%	10%	15%
G. The Universe	15%	10%	8%
H. Energy	15%	10%	10%
I. Motion	10%	10%	10%

The remaining four (J, K, L, and M) highlight scientific applications. These have been adapted and combined to create the reporting category framework for science and technology, shown below.

Content Standard	Application			
	J. Inquiry and Problem Solving	K. Scientific Reasoning	L. Communication	M. Implications of Science & Technology
A. Classifying Life Forms				
B. Ecology				
C. Cells				
D. Continuity and Change				
E. Structure of Matter				
F. The Earth				
G. The Universe				
H. Energy				
I. Motion				

All items in the science and technology assessment measured a content standard; approximately 40% of the items were written to measure a performance indicator in applications.

APPLICATIONS

The score for applications refers to a student's use of knowledge and conceptual and procedural understandings as a basis for application through reasoning, inquiry, communication of ideas, and problem solving.

CONTENT SPECIFICATIONS

The MEA science and technology assessment included multiple-choice, short-answer, and constructed-response items. Each item type was worth a specific number of points in the student's total science and technology score, as shown below.

Type of Item	Possible Score Points
Multiple Choice	0–1
Short Answer	0–2
Constructed Response	0–4

TEST DESIGN

The tables below summarize the numbers and types of items that were used in the MEA science and technology assessment for 2001-02.

GRADE 4

Session	COMMON			MATRIX			Time (minutes)
	MC	SA	CR	MC	SA	CR	
2A	11	2	2	0	0	0	30 (+10)
2B	9	3	2	0	0	0	30 (+10)
2C	0	0	1	8	0	1	30 (+10)

GRADES 8 AND 11

Session	COMMON			MATRIX			Time (minutes)
	MC	SA	CR	MC	SA	CR	
2A	16	5	3	0	0	0	50 (+15)
2B	4	0	2	8	0	1	50 (+15)

Key

- MC = multiple-choice
- SA = short-answer
- CR = constructed-response

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

SCIENCE AND TECHNOLOGY
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common				Matrix			Total Points 194
	MC X 1	SA X 2	CR X 4	Points	MC X 1	CR X 4	Points	
Content	12	3	3	30	58	7	86	116
Classifying Life Forms (Standard A)	2	1		4	2	2	10	14
Ecology (Standard B)	1		1	5	9		9	14
Cells (Standard C)	1		1	5	6	1	10	15
Continuity and Change (Standard D)	3			3	9		9	12
Structure of Matter (Standard E)	2			2	5	1	9	11
The Earth (Standard F)			1	4	6		6	10
The Universe (Standard G)	2	1		4	6	1	10	14
Energy (Standard H)		1		2	8	1	12	14
Motion (Standard I)	1			1	7	1	11	12
Application	8	2	2	20	38	5	58	78
Inquiry and Problem Solving (Standard J)	1		1	5	8	2	16	21
Scientific Reasoning (Standard K)	3	1		5	12	1	16	21
Communication (Standard L)	3	1		5	12	1	16	21
Implications of Science and Technology (Standard M)	1		1	5	6	1	10	15

SCIENCE AND TECHNOLOGY
NUMBER OF POINTS POSSIBLE
GRADE 8

Standard	Common				Matrix			Total Points 194
	MC X 1	SA X 2	CR X 4	Points	MC X 1	CR X 4	Points	
Content	12	3	3	30	57	7	85	115
Classifying Life Forms (Standard A)	3			3	5		5	8
Ecology (Standard B)	3			3	1	1	5	8
Cells (Standard C)	1	1	1	7	10		10	17
Continuity and Change (Standard D)			1	4	10	1	14	18
Structure of Matter (Standard E)	1	1	1	7	3	1	7	14
The Earth (Standard F)				0	7	1	11	11
The Universe (Standard G)	3			3	6	1	10	13
Energy (Standard H)				0	9	1	13	13
Motion (Standard I)	1	1		3	6	1	10	13
Application	8	2	2	20	39	5	59	79
Inquiry and Problem Solving (Standard J)	2		1	6	15	2	23	29
Scientific Reasoning (Standard K)		1	1	6	9		9	15
Communication (Standard L)	3	1		5	12		12	17
Implications of Science and Technology (Standard M)	3			3	3	3	15	18

SCIENCE AND TECHNOLOGY
NUMBER OF POINTS POSSIBLE
GRADE 11

Standard	Common				Matrix			Total Points 194
	MC X 1	SA X 2	CR X 4	Points	MC X 1	CR X 4	Points	
Content	14	2	3	30	64	6	88	118
Classifying Life Forms (Standard A)	3			3	5	1	9	12
Ecology (Standard B)			1	4	6		6	10
Cells (Standard C)	1		1	5	8		8	13
Continuity and Change (Standard D)	1			1	8		8	9
Structure of Matter (Standard E)	2			2	8	1	12	14
The Earth (Standard F)	3	1		5	9	1	13	18
The Universe (Standard G)	2		1	6	5	1	9	15
Energy (Standard H)	2			2	7		7	9
Motion (Standard I)		1		2	8	2	16	18
Application	6	3	2	20	32	6	56	76
Inquiry and Problem Solving (Standard J)	3			3	11	1	15	18
Scientific Reasoning (Standard K)		2	1	8	5		5	13
Communication (Standard L)	3	1		5	9	2	17	22
Implications of Science and Technology (Standard M)			1	4	7	3	19	23

CHAPTER 7—DESIGN OF THE SOCIAL STUDIES ASSESSMENT

BLUEPRINT

The social studies framework was based on Maine's *Learning Results*, which identifies a total of thirteen **content standards** in the four disciplines—civics and government, history, geography, and economics—as listed below:

CIVICS AND GOVERNMENT

- **Rights, responsibilities, and participation:** Students understand the rights and responsibilities of civic life and employ the skills of effective civic participation.
- **Purpose and types of government:** Students understand the types and purposes of governments, their evolution, and their relationships with the governed.
- **Fundamental principles of government and constitutions:** Students understand the constitutional principles and the democratic foundations of the political institutions of the United States.
- **International relations:** Students understand the political relationships among the United States and other nations.

HISTORY

- **Chronology:** Students use the chronology of history and major eras to demonstrate the relationships of events and people.
- **Historical knowledge, concepts, and patterns:** Students develop historical knowledge of major events, people, and enduring themes in the United States, in Maine, and throughout world history.
- **Historical inquiry, analysis, and interpretation:** Students learn to evaluate resource material such as documents, artifacts, maps, artwork, and literature, and to make judgments about the perspectives of the authors and their credibility when interpreting current historical events.

GEOGRAPHY

- **Skills and tools:** Students know how to construct and interpret maps and use globes and other geographic tools to locate and derive information about people, places, regions, and environments.
- **Human interaction with environments:** Students understand and analyze the relationships among people and their physical environments.

ECONOMICS

- **Personal and consumer economics:** Students understand that economic decisions are based on the availability of resources and the costs and benefits of choices.
- **Economic systems of the United States:** Students understand the economic system of the United States, including its principles, development, and institutions.
- **Comparative systems:** Students analyze how different economic systems function and change over time.
- **International trade and global interdependence:** Students understand the patterns and results of international trade.

These thirteen standards have been used to create the reporting category framework for social studies, shown below.

Social Studies Framework		
Standard	**Percentage of Questions Emphasizing Content**	**Percentage of Questions Emphasizing Application**
Civics and Government:		
A. Rights, Responsibilities, and Participation	50%	50%
B./C. Purposes, Types, and Fundamental Principles	60%	40%
D. International Relations	60%	40%
History:		
A./B. Chronology and Historical Knowledge, Concepts, and Patterns	60%	40%
C. Historical Inquiry, Analysis, and Interpretation	40%	60%
Geography:		
A. Skills and Tools	40%	60%
B. Human Interaction with Environments	60%	40%
Economics:		
A. Personal and Consumer Economics	50%	50%
B./C. Economic Systems	50%	50%
D. International Trade and Global Interdependence (Grades 8 and 11)	60%	40%

Social studies education stresses a strong commitment to content knowledge, emphasizes the student's ability to engage in complex thinking and reasoning skills, and emphasizes the clear communication of ideas. Social studies assessment focuses on both content and applications to evaluate what students know and can demonstrate.

CONTENT SPECIFICATIONS

The MEA social studies assessment included multiple-choice, short-answer, and constructed-response items. Each item type was worth a specific number of points in the student's total social studies score, as shown below.

Type of Item	Possible Score Points
Multiple Choice	0–1
Short Answer	0–2
Constructed Response	0–4

TEST DESIGN

The tables below summarize the numbers and item types that were used in the 2001-02 social studies assessment.

GRADE 4

Session	COMMON			MATRIX			Time (minutes)
	MC	SA	CR	MC	SA	CR	
3A	11	2	2	0	0	0	30 (+10)
3B	9	3	2	0	0	0	30 (+10)
3C	0	0	1	8	0	1	30 (+10)

GRADES 8/11

Session	COMMON			MATRIX			Time (minutes)
	MC	SA	CR	MC	SA	CR	
3A	16	5	3	0	0	0	50 (+15)
3B	4	0	2	8	0	1	50 (+15)

Key

- MC = multiple-choice
- SA = short-answer
- CR = constructed-response

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

SOCIAL STUDIES
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common				Matrix			Total Points 194
	MC x 1	SA x 2	CR x 4	Points	MC x 1	CR x 4	Points	
Content	20	5		30	96	1	100	130
Application			5	20		11	44	64
Civics and Government (Standards A, B, and C)	8	1	1	14	23	1	27	41
Rights, Responsibilities, and Participation (Standard A)	4		1	8	9		9	17
Purpose, Types, and Fundamental Principles of Government (Standards B and C)	4	1		6	14	1	18	24
History (Standards A, B, and C)	3	1	2	13	26	4	42	55
Chronology, Historical Knowledge, Concepts, and Patterns (Standards A and B)	2	1	1	8	25	2	33	41
Historical Inquiry, Analysis, and Interpretation (Standard C)	1		1	5	1	2	9	14
Geography (Standards A and B)	7	2	1	15	27	4	43	58
Skills and Tools (Standard A)	2	1		4	17	3	29	33
Human Interaction with Environments (Standard B)	5	1	1	11	10	1	14	25
Economics (Standards A, B, C, and D)	2	1	1	8	20	3	32	40
Personal and Consumer Economics/ Economic Systems (Standards A and B)	2	1	1	8	17	3	29	37
Comparative Systems/International Trade and Global Interdependence (Standards C and D)				0	3		3	3

SOCIAL STUDIES
NUMBER OF POINTS POSSIBLE
GRADE 8

Standard	Common				Matrix			Total Points 194
	MC x 1	SA x 2	CR x 4	Points	MC x 1	CR x 4	Points	
Content	20	4		28	88	1	92	120
Application		1	5	22	8	11	52	74
Civics and Government (Standards A, B, C and D)	5	2	1	13	26	2	34	47
Rights, Responsibilities, and Participation (Standard A)	2			2	8	1	12	14
Purpose, Types, and Fundamental Principles of Government (Standards B and C)	2	1	1	8	14		14	22
International Relations (Standard D)	1	1		3	4	1	8	11
History (Standards A, B, and C)	6		2	14	26	5	46	60
Chronology, Historical Knowledge, Concepts, and Patterns (Standards A and B)	6		2	14	22	2	30	44
Historical Inquiry, Analysis, and Interpretation (Standard C)				0	4	3	16	16
Geography (Standards A and B)	5	2	1	13	22	4	38	51
Skills and Tools (Standard A)	3	1		5	11	1	15	20
Human Interaction with Environments (Standard B)	2	1	1	8	11	3	23	31
Economics (Standards A, B, C, and D)	4	1	1	10	22	1	26	36
Personal and Consumer Economics (Standards A)	1			1	8		8	9
Economic Systems/Comparative Systems (Standards B and C)	3			3	11	1	15	18
International Trade and Global Interdependence (Standards D)		1	1	6	3		3	9

SOCIAL STUDIES
NUMBER OF POINTS POSSIBLE
GRADE 11

Standard	Common				Matrix			Total Points 194
	MC x 1	SA x 2	CR x 4	Points	MC x 1	CR x 4	Points	
Content	13	3		19	78		78	97
Application	7	2	5	31	18	12	66	97
Civics and Government (Standards A, B, C, and D)	5	2	1	13	26	3	38	51
Rights, Responsibilities, and Participation (Standard A)				0	7	2	15	15
Purpose, Types, and Fundamental Principles of Government (Standards B and C)	3	2	1	11	15	1	19	30
International Relations (Standard D)	2			2	4		4	6
History (Standards A, B, and C)	6		2	14	25	4	41	55
Chronology, Historical Knowledge, Concepts, and Patterns (Standards A and B)	6		1	10	23	3	35	45
Historical Inquiry, Analysis, and Interpretation (Standard C)			1	4	2	1	6	10
Geography (Standards A and B)	5	2	1	13	24	3	36	49
Skills and Tools (Standard A)	2	1	1	8	13	2	21	29
Human Interaction with Environments (Standard B)	3	1		5	11	1	15	20
Economics (Standards A, B, and D)	4	1	1	10	21	2	29	39
Personal and Consumer Economics (Standards A)				0	6	2	14	14
Economic Systems/Comparative Systems (Standards B and C)	3			3	11		11	14
International Trade and Global Interdependence (Standard D)	1	1	1	7	4		4	11

CHAPTER 8—DESIGN OF THE VISUAL AND PERFORMING ARTS ASSESSMENT

BLUEPRINT

The visual and performing arts assessment includes four disciplines: dance, music, theater, and visual arts. The arts framework is based on Maine's *Learning Results*, which identifies three content standards in the arts as listed below:

- **Creative expression:** Students create and/or perform to express ideas and feelings.
- **Cultural heritage:** Students understand the cultural contributions (social, ethical, political, religious dimensions) of the arts, how the arts shape and are shaped by prevailing cultural and social beliefs and values, and recognize exemplary works from a variety of cultures and historical periods.
- **Criticism and aesthetics:** Students reflect upon and assess the characteristics and merits of art works.

These three standards were used to create the reporting category framework for the visual and performing arts, as shown below.

Visual and Performing Arts Framework

Discipline	Standard		
	A. Creative Expression	B. Cultural Heritage	C. Criticism and Aesthetics
Dance			
Music			
Theater			
Visual Arts			

Each row and each column of the framework constitutes a reporting category for school- and district-level results in the MEA—for example, music/cultural heritage. Student-level results were not reported in the visual and performing arts, as no common items were used in this area.

It should be noted that not all of the performance indicators associated with each content standard (see *Learning Results*) can be assessed reliably and validly using a paper-and-pencil test. For example, some of the

performance indicators included under the standard for "creative expression" would best be measured in other ways. For this reason, additional methods of assessment for these performance indicators are being studied.

The distribution of items, or emphasis, across the arts disciplines in the MEA varies from one grade level to another, as shown in the table below.

Discipline	Grade		
	4	8	11
Dance	10%	10%	15%
Music	40%	40%	35%
Theater	10%	10%	15%
Visual Arts	40%	40%	35%

CONTENT SPECIFICATIONS

The MEA visual and performing arts assessment included multiple-choice and constructed-response items. Each item type was worth a specific number of points, as shown below:

Type of Question	Possible Score Points
Multiple Choice	0–1
Constructed Response	0–4

TEST DESIGN

The table below summarizes the numbers and types of matrix-sampled items that were used in the 2001-02 visual and performing arts assessment.

Visual and Performing Arts

Session	MATRIX			Time (minutes)
	MC	SA	CR	
5A	7	0	2	25 (+10)

Key

- MC = multiple-choice
- CR = constructed-response

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

VISUAL AND PERFORMING ARTS
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common			Matrix			Total Points 120
	MC	CR	Points	MC x 1	CR x 4	Points	
Dance				16	3	28	28
Music				26	3	38	38
Theater				13	3	25	25
Visual Arts				17	3	29	29
Creative Expression (Standard A)				28	4	44	44
Cultural Heritage (Standard B)				21	3	33	33
Criticism and Aesthetics (Standard C)				23	5	43	43

VISUAL AND PERFORMING ARTS
NUMBER OF POINTS POSSIBLE
GRADE 8

Standard	Common			Matrix			Total Points 120
	MC	CR	Points	MC x 1	CR x 4	Points	
Dance				9	3	21	21
Music				26	3	38	38
Theater				9	3	21	21
Visual Arts				28	3	40	40
Creative Expression (Standard A)				30	3	42	42
Cultural Heritage (Standard B)				19	5	39	39
Criticism and Aesthetics (Standard C)				23	4	39	39

VISUAL AND PERFORMING ARTS
NUMBER OF POINTS POSSIBLE
GRADE 11

Standard	Common			Matrix			Total Points 120
	MC	CR	Points	MC x 1	CR x 4	Points	
Dance				11	3	23	23
Music				27	2	35	35
Theater				11	3	23	23
Visual Arts				23	4	39	39
Creative Expression (Standard A)				28	4	44	44
Cultural Heritage (Standard B)				20	4	36	36
Criticism and Aesthetics (Standard C)				24	4	40	40

CHAPTER 9—DESIGN OF THE HEALTH EDUCATION ASSESSMENT

BLUEPRINT

The health framework was based on Maine's *Learning Results*, which identifies six **content standards** as shown below:

- **Health concepts:** Students understand health promotion and disease prevention concepts.
- **Health information, services, and products:** Students know how to acquire valid information about health issues, services, and products.
- **Health promotion and risk reduction:** Students understand how to reduce their health risks through the practice of healthy behaviors.
- **Influences on health:** Students understand how media techniques, cultural perspectives, technology, peers, and family influence behaviors that affect health.
- **Communication skills:** Students understand that skillful communication can contribute to better health for them, their families, and the community.
- **Decision making and goal setting:** Students learn how to set personal goals and make decisions that lead to better health.

These six standards were combined with the ten health education content areas identified by the 1984 Education Reform Act to create a reporting category framework for health, as shown on the next page.

Health Framework						
	Health Standard					
Content Area	A. Health Concepts	B. Health Information, Services, and Products	C. Health Promotion and Risk Reduction	D. Influences on Health	E. Communication Skills	F. Decision Making and Goal Setting
Community, Consumer, and Environmental Health						
Personal and Nutritional Health						
Family Life Education and Growth and Development						
Safety and Injury Prevention						
Tobacco, Alcohol, and Other Drug Use Prevention						
Prevention and Control of Disease and Disorders						
Total	30%	70%				

Thirty percent of the items measured health standard A; they were divided among the six content areas. The remaining 70% of the items was divided among B through F and the six content areas. The distribution of items was 10% to 20% for each standard, determined by its developmental appropriateness for the specific grade being assessed.

A portion of the items in the health assessment were developed by the Health Education Assessment Project for the State Collaborative on Assessment and Student Standards (SCASS) under the auspices of the Council of Chief State School Officers. Each SCASS item that was used or adapted was aligned with a performance indicator from Maine's health education standards. Maine educators on the content development committee developed the remainder of the items.

CONTENT SPECIFICATIONS

The MEA health assessment included multiple-choice, short-answer, and constructed-response items. Each item type was worth a specific number of points in the student's total health score, as shown below.

Type of Item	Possible Score Points
Multiple Choice	0–1
Short Answer	0–2
Constructed Response	0–4

TEST DESIGN

At every grade level, the assessment included no common items but was constructed solely of matrix-sampled items. The table below summarizes the numbers and item types that were used in the 2001-02 health education assessment for all grade.

GRADES 4, 8, AND 11

Session	MATRIX			Time (minutes)
	MC	SA	CR	
5A	8	0	3	40 (+10)

Key

- MC = multiple-choice
- CR = constructed-response

The charts on the following pages outline the total number of possible points—as reported—by learning results and item type.

HEALTH EDUCATION
NUMBER OF POINTS POSSIBLE
GRADE 4

Standard	Common		Matrix		Total Points 223
	MC	CR	MC X 1	CR X 4	
Health Concepts (Standard A)			33	7	61
Health Information, Services, and Products (Standard B)			18	4	34
Health Promotion and Risk Reduction (Standard C)			21	3	33
Influences on Health (Standard D)			15	2	23
Communication Skills (Standard E)			1	11	45
Decision Making and Goal Setting (Standard F)			7	5	27
Community, Consumer, and Environmental Health			15	6	39
Personal and Nutritional Health			26	7	54
Family Life Education and Growth and Development			14	4	30
Safety and Injury Prevention			17	6	41
Tobacco, Alcohol, and Other Drug Use Prevention			16	6	40
Prevention and Control of Disease and Disorders			7	3	19

HEALTH EDUCATION
NUMBER OF POINTS POSSIBLE
GRADE 8

Standard	Common		Matrix		Total Points 224
	MC	CR	MC X 1	CR X 4	
Health Concepts (Standard A)			56	6	80
Health Information, Services, and Products (Standard B)			9	3	21
Health Promotion and Risk Reduction (Standard C)			12	9	48
Influences on Health (Standard D)			8	5	28
Communication Skills (Standard E)			3	4	19
Decision Making and Goal Setting (Standard F)			8	5	28
Community, Consumer, and Environmental Health			15	5	35
Personal and Nutritional Health			12	7	40
Family Life Education and Growth and Development			17	6	41
Safety and Injury Prevention			18	6	42
Tobacco, Alcohol, and Other Drug Use Prevention			23	6	47
Prevention and Control of Disease and Disorders			11	2	19

HEALTH EDUCATION
NUMBER OF POINTS POSSIBLE
GRADE 11

Standard	Common		Matrix		Total Points 224
	MC	CR	MC X 1	CR X 4	
Health Concepts (Standard A)			38	7	66
Health Information, Services, and Products (Standard B)			9	3	21
Health Promotion and Risk Reduction (Standard C)			19	7	47
Influences on Health (Standard D)			14	4	30
Communication Skills (Standard E)			11	6	35
Decision Making and Goal Setting (Standard F)			5	5	25
Community, Consumer, and Environmental Health			15	2	23
Personal and Nutritional Health			28	3	40
Family Life Education and Growth and Development			14	6	38
Safety and Injury Prevention			17	6	41
Tobacco, Alcohol, and Other Drug Use Prevention			12	8	44
Prevention and Control of Disease and Disorders			10	7	38

SECTION II: TEST ADMINISTRATION

CHAPTER 10—TEST ADMINISTRATION

RESPONSIBILITY FOR ADMINISTRATION

As indicated in the *Principal/Test Coordinator's Manual*, principals and/or their designated MEA coordinator were responsible for the proper administration of the MEA. Manuals and certification forms were used to ensure the uniformity of administration procedures from school to school.

PROCEDURES

Principals and/or the school's designated MEA coordinator were instructed to read the *Principal/Test Coordinator's Manual* prior to testing and to be familiar with the instructions given in the *Test Administrator's Manual*. The *Principal/Test Coordinator's Manual* provided each school with checklists to help them to prepare for testing. The checklists outlined tasks for the schools to perform before, during, and after test administration. Along with these checklists, the *Principal/Test Coordinator's Manual* outlined the nature of the testing material being sent to each school, how to inventory the material, how to track it during administration, and how to return the material once testing was complete. It also contained information about including or excluding students. The *Test Administrator's Manual* also included checklists for the administrators to prepare themselves, their classrooms, and the students for the administration of the test. The *Test Administrator's Manual* contained sections that detailed the procedures to be followed for each test session, and it contained instructions on preparing the material prior to giving it to the principal/coordinator for its return to Measured Progress.

ADMINISTRATOR TRAINING

In addition to distributing the *Principal/Test Coordinator's* and *Test Administrator's Manual*s, the Maine Department of Education, along with Measured Progress, conducted four test administration workshops to train and inform school personnel about the MEA. Three live workshops were presented in Bangor, Presque Isle, and Portland, and one workshop was broadcast over the state's closed-circuit television network.

PARTICIPATION REQUIREMENTS

All students who were considered for accommodations on the MEA were to have had their individual situations reviewed by a group within the school prior to the time of testing. For every student with an identified exceptionality requiring an Individual Educational Plan (IEP), schools were required to hold a Pupil Evaluation Team (PET) meeting that addressed that student's needs for modifications. For other students needing test accommodations who did not have an identified exceptionality, a meeting was required that included one of the student's teachers, the building principal, related services personnel, and, whenever possible, the student's parents. If it was not possible for the parents to attend the meeting, it was required that they be notified of the committee's recommendations for accommodations prior to the time of testing.

Recommended accommodations were to be consistent with those accommodations already being employed in the student's instructional program. Any such accommodations were reflected either in the minutes of the PET meeting (for students requiring an IEP) or in a statement prepared for the cumulative folders of students not requiring IEPs. The following is the suggested statement that schools were given as a model:
The student will/will not participate in the __th-grade Maine Educational Assessment as scheduled during the month of _____ 19__. The following test accommodations will be observed: (list accommodations)

EXCLUSION FROM THE ASSESSMENT

The legislation's intent is for **all** students in grades 4, 8, and 11 to participate in the MEA through standard administration, administration with accommodations, or alternate assessment. Furthermore, any student who is absent during any session or sessions of the MEA is expected to take a makeup test within the two-week testing window. Exclusion was to be considered only as a last resort.

On those occasions where it was deemed necessary to exclude a student from sections of the assessment or from the assessment as a whole, schools were asked to seek the approval of the Department of Education. It was recommended that the exclusion be limited to only those sections of the MEA that were considered inappropriate for that particular student. Exclusion was to be selected only after the various types of modifications available had been fully explored, and it was felt that the assessment would not yield a valid indication of how a student functioned in a given content area. For example, even students who were reading two years below grade level were advised to take

the reading section because those scores would give a fair representation of their current level of functioning in reading. If, however, after examining all of the possible modifications, a local school decided that the assessment or sections of it would be inappropriate for a given student, that student could be excluded.

DOCUMENTATION OF MODIFICATIONS OR EXCLUSIONS

Information about the modifications given to students or the reasons for exclusion was provided on the front page of the student's response booklet. This information was coded in by staff, not students, after testing was completed. The *Principal/Test Coordinator's* and *Test Administrator's Manual* provided directions on coding in the information related to modification(s), partial exclusion, and exclusion, and every student who was totally excluded had to be accounted for in the designated section of the response booklet.

STATE PARTICIPATION RATES—FALL 2001

Student Participation Category	Number			Percentage		
	Gr. 4	Gr. 8	Gr. 11	Gr. 4	Gr. 8	Gr. 11
Students enrolled on the first day of testing	16168	17488	15644	100	100	100
Ethnicity	16168	17488	15644	100	100	100
White (non-Hispanic)	14830	15912	14280	92	91	91
Black (non-Hispanic)	189	201	114	1	1	1
Hispanic	86	143	128	1	1	1
Asian/Pacific Islander	168	171	156	1	1	1
American Indian/Alaskan Native	182	224	159	1	1	1
Multi-ethnic	249	385	233	2	2	1
Other	144	232	214	1	1	1
Not reported	320	220	360	2	1	2
Internet access at home	16168	17488	15644	100	100	100
Yes	10241	13636	12441	63	78	80
No	5927	3852	3203	37	22	20
Students who took all or part of the assessment without accommodations	13469	15105	13827	83	86	88
Students who took all or part of the assessment with accommodations	2383	1876	1104	15	11	7
Identified disability (PET/IEP)	1747	1568	963	73	84	87
LEP	39	18	1	2	1	0
504 Plan	55	28	27	2	1	2
Other	329	5	2	14	0	0
Reason not reported	237	258	111	10	14	10
Students recommended for participation in alternate assessment (PAAP)	196	210	133	1	1	1
Identified disability (PET/IEP)	165	173	123	84	82	92
LEP	28	31	7	14	15	5
504 Plan	0	0	0	0	0	0
Reason not reported	3	6	3	2	3	2
Students who did not participate in all or part of the assessment due to absence	78	226	486	0	1	3
Students who did not participate in all or part of the assessment due to other reasons	42	71	94	0	0	1

STATE PARTICIPATION RATES—SPRING 2002

Student Participation Category	Number			Percentage		
	Gr. 4	Gr. 8	Gr. 11	Gr. 4	Gr. 8	Gr. 11
Students enrolled on the first day of testing	16198	17445	15315	100	100	100
Ethnicity	16198	17445	15315	100	100	100
White (non-Hispanic)	14956	15949	13951	92	91	91
Black (non-Hispanic)	192	184	135	1	1	1
Hispanic	112	154	127	1	1	1
Asian/Pacific Islander	174	185	159	1	1	1
American Indian/Alaskan Native	199	242	147	1	1	1
Multi-ethnic	219	309	209	1	2	1
Other	133	264	217	1	2	1
Not reported	213	158	370	1	1	2
Internet access at home	16198	17445	15315	100	100	100
Yes	11358	13847	12067	70	79	79
No	4840	3598	3248	30	21	21
Students who took all or part of the assessment without accommodations	13416	14962	13579	83	86	89
Students who took all or part of the assessment with accommodations	2437	1973	1076	15	11	7
Identified disability (PET/IEP)	1749	1606	927	72	81	86
LEP	50	52	7	2	3	1
504 Plan	70	24	30	3	1	3
Other	294	10	3	12	1	0
Reason not reported	304	285	111	12	14	10
Students recommended for participation in alternate assessment (PAAP)	194	178	82	1	1	1
Identified disability (PET/IEP)	167	150	71	86	84	87
LEP	21	21	7	11	12	9
504 Plan	0	0	0	0	0	0
Reason not reported	7	7	4	4	4	5
Students who did not participate in all or part of the assessment due to absence	117	294	514	1	2	3
Students who did not participate in all or part of the assessment due to other reasons	34	38	64	0	0	0

TESTING IRREGULARITIES

There were **no** testing irregularities for the 2001–02 assessment year.

SECTION III: DEVELOPMENT AND REPORTING OF SCORES

CHAPTER 11—SCORING

MACHINE SCORED ITEMS

Once the 2001-02 booklets had been logged in, identified with appropriate scannable, pre-printed school information sheets, examined for extraneous materials, and batched, they were moved into the scanning area. For all response booklets (and questionnaires and other forms that require imaging/scanning) to be imaged, this area is the last stop in the processing loop in which the documents themselves are handled.

At that point, 100% of the response document and other scannable information necessary to produce the required reports had been captured and converted into an electronic format, including all student identification and demographics, selected-response answers, and digital image clips of hand-written responses. The digital image clip information allowed Measured Progress to replicate student responses just as they appeared on the originals, but they had been transferred onto the readers' monitors. From that point on, the entire process—data processing, scoring, "range-finding," data analysis, reporting—was accomplished without further reference to the originals.

The first step in that conversion was the removal of the booklet bindings so that the individual pages could pass through the scanners, one at a time. Once cut, the sheets were put back in their proper boxes and placed in storage until needed for the scanning/imaging process.

Customized scanning programs for all scannables were prepared to selectively read the student response booklets and to format the scanned information electronically according to pre-determined requirements. Any information (including multiple-choice response data) that had been designated time-critical or process-critical was handled first.

In addition to numerous real-time quality control checks, duplex read, a transport printer that prints a unique identifying number on each sheet of each booklet, and on-line editing capability, the 5000i scanners offer features that make them compatible with Internet technology.

SCANNING QUALITY CONTROL

NCS scanners are equipped with many built-in safeguards that prevent data errors. The scanning hardware is continually monitored for conditions that will cause the machine to shut down if standards are not met. It will display an error message and prevent further scanning until the condition is corrected. The areas monitored include document page and integrity checks, user-designed on-line edits, and many internal checks of electronic functions.

Before every scanning shift begins, Measured Progress's operators performed a daily diagnostic routine. This is yet another step to protect data integrity, and one that has been done faithfully for the many years that we have been involved in production scanning. In the rare event that the routine detects a photocell that appears to be out of range, we calibrate that machine and perform the test again. If the read is still not up to standard, we call for assistance from our field service engineer.

As a final safeguard, spot checks of scanned files, bubble by bubble and image by image, were routinely made throughout scanning runs. The result of these precautions, from the original layout of the scanning form to the daily vigilance of our operators, was a scan error rate well below 0.001.

ELECTRONIC DATA FILES

Once the data had been entered and the scanning logs and other paperwork completed, the booklets themselves were put into storage (where they stayed for at least 180 days beyond the close of the fiscal year). When it had been determined that the files were complete and accurate, those files were duplicated electronically and made available for many other processing options. Completed files were loaded onto our local area network (LAN) for transfer to Measured Progress' proprietary I-Score system for scoring. Those files were then used to identify (and print out) papers to be used in the rangefinding and standard-setting processes and the data was made transferable via the Internet, CD-ROM, or optical disk.

ITEMS SCORED BY READERS

Test and answer materials were handled as little as possible to minimize the possibility of loss, mishandling, or breach of security. Once scanned, either by optical mark reader or the I-Score system, papers were stored securely in areas with limited personnel access.

As explained in the following sections on scoring, the I-Score system itself ensures the security of responses and test items: all scoring is "blind"; that is, no student names are associated with viewed responses or raw scores and all scoring personnel are subject to the same nondisclosure requirements and supervision as regular Measured Progress staff.

I-SCORE

After the 2001-02 test material had been loaded into the LAN, I-Score sent electronically scanned images of student work to individual readers at computer terminals who evaluated each response and recorded each student's score via keypad or mouse entry. When the reader had finished with one response, the next response appeared immediately on the computer screen. In that way, the system guaranteed complete anonymity of individual students and ensured the randomization of responses during scoring.

Although I-Score is based on conventional scoring techniques, it also offers numerous benefits, not the least of which is raising the bar on scoring process capability. Some of the benefits are as follows:

- real-time information on scorer reliability, read-behinds, and overall process monitoring;
- early access to subsets of data for tasks such as standard setting;
- reduced material handling, which not only saves time and labor, but also enhances the security of materials; and
- immediate access to samples of student responses and scores for reporting and analysis through electronic media.

Scoring operations, directed by the manager of scoring services, are carried out by a highly qualified staff. The staff included:

- chief readers, who oversaw all training and scoring within particular subject areas;
- quality assurance coordinators (QACs), who lead rangefinding and training activities and monitor scoring consistency and rates;
- verifiers, who perform read-behinds of readers and assist at scoring tables as necessary; and
- readers, who perform the bulk of the scoring.

Table 11-1 summarizes the qualifications of the 2001-02 MEA quality assurance coordinators and readers.

Table 11-1
Qualifications of 2001-02 QACs and Readers

2001 Fall Administration					
Scoring Responsibility	Educational Credentials				Total
	Doctorate	Masters	Bachelors	Other	
QACs	6.25	37.5	43.75	12.5	100%
Readers	3.91	28.91	53.13	14.06	100%
2002 Spring Administration					
Scoring Responsibility	Educational Credentials				Total
	Doctorate	Masters	Bachelors	Other	
QACs	7.41	48.15	33.33	11.11	100%
Readers	.95	17.62	62.38	19.04	100%

PRELIMINARY ACTIVITIES

Preliminary activities for scoring included (1) participating in the planning and design of documents to be used for scoring, (2) reviewing items and score guides for rangefinding and training and the creation of rangefinding packets, and (3) selecting scoring staff and training them for scoring.

PLANNING AND DESIGNING DOCUMENTS

Scoring personnel advised project management and DOE staff on the program design in order to support an efficient and effective scoring process. Scoring staff contributed also to the design of

- response documents and the image-capture process to yield acceptable image clips (also defining file format and layout); and
- scoring benchmarks composed of the guide, subject background information, and anchor papers.

REVIEWING ITEMS AND GUIDES (RANGEFINDING)

Before the scheduled start of scoring activities, scoring center staff reviewed test items and scoring guides for rangefinding. At that point, chief readers and selected QACs prepared scorer training materials. Measured Progress's scoring staff (including test developers) selected one or two anchor examples for each item score point. An additional six to ten responses per item were chosen as part of the training pack. The anchor pack consisted of mid-range exemplars, while the training pack exemplars illustrated the range within each score point. The chief readers, who worked closely with QACs for each content area, facilitated the selection of response exemplars. One of the greatest difficulties in the selection of anchor and training exemplars was finding a sufficient number of papers representing the highest scores (4 or 8) as such scores are fairly rare.

SELECTING AND TRAINING SCORING STAFF

SELECTING QUALITY ASSURANCE COORDINATORS (QACS) AND VERIFIERS

Because the read-behinds performed by the QACs and verifiers moderated the scoring process and thus maintained the integrity of the scores, individuals to fill those positions were selected for their accuracy. In addition, QACs, who train readers to score each item in their content areas, were selected for their ability to instruct and for their level of expertise in their content areas. For this reason, QACs typically are retired teachers who have demonstrated a high level of expertise in their respective disciplines. The ratio of QACs and verifiers to readers was approximately 1:11.

TRAINING QUALITY ASSURANCE COORDINATORS AND VERIFIERS

To ensure that all QACs provided consistent training and feedback, the chief readers spent two days training and qualifying the QACs, and the QACs reviewed all items with the verifiers before scoring. In addition, QACs rotated among tables, supervising readers and reading behind verifiers, who in turn read behind a different table of readers each day.

SELECTING READERS

Applicants were required to demonstrate their ability by participating in a preliminary scoring evaluation. The I-Score system enables Measured Progress to efficiently measure a prospective reader's ability to score student responses accurately. After having participated in a training session, applicants were required to achieve at least 80% exact scoring agreement for a qualifying pack consisting of 20 responses to a predetermined item in their content area. Those 20 responses were randomly selected from a bank of approximately 150, all of which had been selected by QACs and approved by the chief readers and developers.

TRAINING READERS

The QACs first applied the language of the scoring guide for an item to its anchor pack exemplars. Once discussion of the anchor pack had concluded, readers attempted to score the training pack exemplars correctly. The QACs then reviewed the training pack and answered any questions readers had before actual scoring began. With this system, two aspects of scoring efficiency are in conflict. First, in order to minimize training expense, it is desirable to train each reader on as few items as possible. Second, to prevent reader drift and to minimize retraining

requirements, it is desirable to score a given item in a brief period of time. But the lower the number of unique items each reader scores, the greater the number of readers required to score that item quickly. To minimize that conflict, we divided each subject area's readers into two or more groups. On the first day of scoring, each group was trained to score a different item. When a group had completed all of an item's responses, those readers were trained on another item (or set).

SCORING ACTIVITIES

Student response booklets were digitally scanned and scored on a file server for a dedicated, secure LAN. I-Score then distributed digital images of student responses to readers. Training and scoring took place over a period of approximately two weeks. Items were randomly assigned to readers; thus, each item in a student's response booklet was more than likely scored by a different reader. By using the maximum possible number of readers for each student, the procedure effectively minimized error variance due to reader sampling. All common and matrix constructed-response items were scored once with a 2% read-behind to ensure consistency among readers and accuracy of individual readers.

MONITORING READERS

After a reader scored a student response, I-Score determined whether that response should also be scored by another reader, scored by a QAC or verifier, or routed for special attention. QACs and verifiers used I-Score to produce daily reader accuracy and speed reports. QACs and verifiers were able to obtain current reader accuracy reports and speed reports on-line at any time.

SCORING THE WRITING

Maine teachers and administrators were recruited to score the common writing prompt at in-state scoring sessions that were held in Bangor and Gorham, Maine. Teachers who participated in the scoring process developed skills in holistic evaluation of writing using a rubric aligned with the standards outlined in the Maine *Learning Results*. Those skills could then be applied to writing instruction in the classrooms, and the scoring of writing also gave participants an opportunity to read the range of student writing produced at each grade and to connect their current teaching practices with the recommendations in the Maine *Learning Results*. Administrators who participated gained skills helpful in improving the teaching and evaluation of writing in their schools. Maine teachers' involvement in scoring also created a

network of teachers who served as a resource to their local and state schools. Beginning with the 2001-02 MEA, use of annotations in the scoring of writing was discontinued.

General Scoring Guides

Short-Answer Items

Score Point	Description
2	▪ The student's response provides a complete and correct answer.
1	▪ The student's response is partially correct. ▪ The student's response may be incomplete or contain errors.
0	▪ The student's response is totally incorrect or too minimal to evaluate.
B	▪ Blank/no response.

Constructed-response Items

Score Point	Description
4	▪ The student completes all important components of the task and communicates ideas clearly. ▪ The student demonstrates in-depth understanding of the relevant concepts and/or processes. ▪ When instructed to do so, the student chooses more efficient and/or sophisticated processes. ▪ When instructed to do so, the student offers insightful interpretations or extensions (e.g., generalizations, applications, and analogies).
3	▪ The student completes the most important components of the task and communicates clearly. ▪ The student demonstrates understanding of major concepts even though he/she overlooks or misunderstands some less important ideas or details.
2	▪ The student completes most important components of the task and communicates those clearly. ▪ The student demonstrates that there are gaps in his/her conceptual understanding.
1	▪ The student shows minimal understanding. ▪ The student addresses only a small portion of the required task(s).
0	▪ The student's response is totally incorrect or irrelevant.
B	▪ Blank/no response.

MEA WRITING SCORING GUIDE 2001-02

Stylistic & Rhetorical Aspects of Writing Topic Idea Development					
1	**2**	**3**	**4**	**5**	**6**
▪ Little topic development and/or organization, few details ▪ Possible evidence of voice ▪ Simplistic language (wording and sentence structures)	▪ Limited topic development, focus, and/or details ▪ Evidence of voice ▪ Limited variety in language used (wording and sentence structures)	▪ Moderate topic development, focus, and details ▪ Some voice ▪ Some variety in language used (wording and sentence structures)	▪ Well developed with control and relevant details ▪ Consistent voice ▪ Variety in language used (wording and sentence structures)	▪ Fully developed with strong details ▪ Sustained voice and/or tone with emerging style ▪ Effective use of language	▪ Topic and details richly developed ▪ Distinctive voice, tone and style ▪ Rich use of language

Topic Development	The overall effect of the paper
Organization	The degree to which the response is: ▪ Focused ▪ Clearly and logically ordered ▪ Clarified by paragraphs
Details	The degree to which the response includes examples that develop the main points.
Language/Style	The degree to which manipulation of language, including vocabulary, word choice, word combination, and sentence variety is effective

Standard English Conventions			
1	**2**	**3**	**4**
▪ Errors seriously interfere with communication and/or ▪ Little control of sentence structure, grammar and usage, and mechanics in **first draft writing**	▪ Errors interfere somewhat with communication and/or ▪ Few or no errors in simplistic or limited text in **first draft writing**	▪ Errors do not interfere with communication and/or ▪ Few errors relative to length of essay or complexity of sentence structure, grammar and usage, and mechanics in **first draft writing**	▪ Control of a variety of sentence structures, grammar and usage, and mechanics ▪ Length and complexity of essay provide opportunity for student to show control of standard English conventions in **first draft writing**

Sentences	The degree to which the response includes sentences that are correct in structure
Grammar and Usage	The degree to which the response demonstrates correct ▪ Use of standard grammatical rules of English ▪ Word usage and vocabulary
Mechanics	The degree to which the response demonstrates correct ▪ Punctuation ▪ Capitalization ▪ Spelling

CHAPTER 12: EQUATING AND SCALING

Scaled scores for the 2001-02 MEA reading, writing, mathematics, science and technology, and social studies tests were developed by equating the 2001-02 scores to the 2000-01 scores. Equating the scores from alternate forms of a test adjusts for any difference in difficulty and ensures that scores from the different forms are comparable. Because the 2000-01 and 2001-02 versions of each test are developed from the same framework, they may be considered alternate forms. Equating test scores from the 2000-01 and 2001-02 administrations of each test makes it possible to report the results of the 2001-02 administration on the same scale used in the previous year. The equated scores then get transformed to scaled scores. The process of equating and scaling does not change the rank ordering of students, give more weight to particular questions, or change students' performance level classifications.

Equating for MEA used the *anchor-test-nonequivalent-groups* design with external anchor described by Petersen, Kolen, & Hoover (1993). The "anchor test" for reading, mathematics, science and technology, and social studies is a set of matrix items included in both test administrations. These items are external to the test in that they do not contribute to the students' raw scores in either administration of the test. For writing, the reading test was used as the "anchor test." Because reading scores for 2000-01 and 2001-02 were equated, the reading scores for the two years are equivalent and can be used in the same way as a set of common items.

The students who took a given test in 2000-01 and 2001-02 are naturally occurring groups, so no assumption could be made regarding their equivalence. Item Response Theory (IRT) is particularly useful in equating for nonequivalent groups (Allen & Yen, 1979). All IRT calibrations performed on the MEA are used for equating purposes only.

Developing equated scores for the 2001-02 MEA involved several steps. The first step was to construct the "anchor test;" that is, to determine the set of equating items. This step did not apply in the case of the writing test. The second step was to calibrate the items in an IRT model. In the item calibration process, the two "forms" of the test (2000-01 and 2001-02) were calibrated to the same score scale using the anchor test. Finally, in the third step, raw score cutpoints were determined for the 2000-01-02 test and scaling transformation constants were calculated. These values were used to compute the scaled scores, which were then used to report the MEA results.

DETERMINING THE SETS OF EQUATING ITEMS

During the development stage of MEA 2001-02, matrix items that were also administered in 2000-01 were identified as potential equating items. These items were designated based on the following criteria:

1. The average difficulty of the equating items was about the same as their average difficulty on the 2000-01 test.
2. The total points from the equating items are about equivalent to 40% of the total points on the test.
3. The position of each item in the 2001-02 form was about the same as its position in the 2000-01 form.
4. The distribution of the items across different relevant categories (i.e. item types and content areas) was similar to that of the whole test.
5. There was not any significant change in the item from one administration to the other.

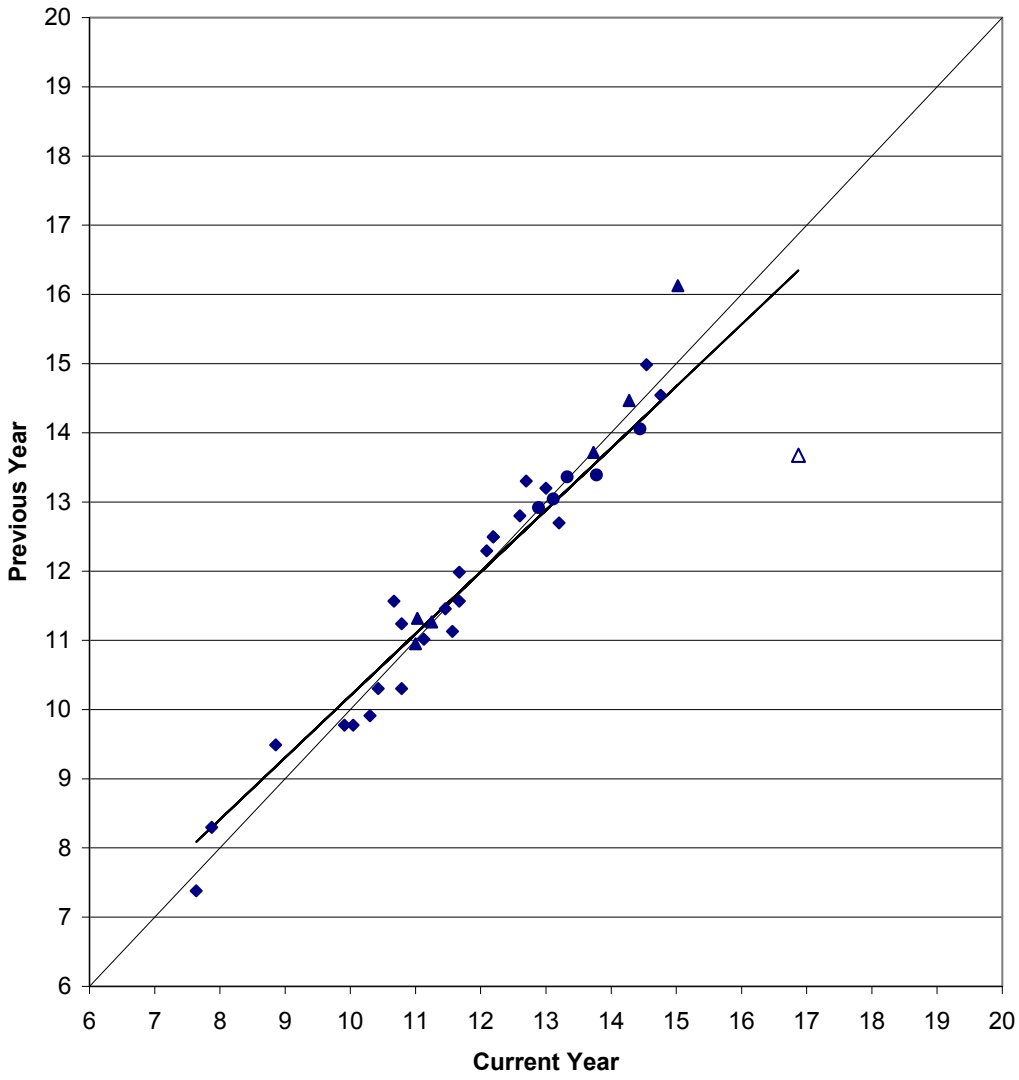
To determine the final set of equating items for each grade level and subject combination, a differential item functioning (DIF) approach using the delta plot method was applied. The p-values of each multiple-choice and short-answer item were transformed to the delta metric. Each item has two p-values, one for each test administration. The delta scale is an inverse normal transformation of

percentage correct to a linear scale with a mean of 13 and standard deviation of 4 (Holland & Wainer, 1993). A high delta value indicates a difficult item. For constructed-response items, the adjusted p-value (the average score divided by the maximum possible score) was transformed to the delta metric. The delta values computed for the potential equating items were plotted for each subject (reading, mathematics, science and technology, social studies, health and VPA) in each grade level (4, 8, 11).

Figure 12-1 is an example of delta plot for equating items. The dark diagonal line is the trend line and the light diagonal line is the identity line. Different shapes were used to identify different item types: ♦ for multiple-choice items; ▲ for short-answer items; and, ● for constructed-response items. The perpendicular distance of each item to the regression line was computed. The unshaded shape indicates the item with the greatest perpendicular distance from the regression line. Items that were not more than three standard deviations away from the regression line were used as equating items. The delta plots are included in Appendix C.

An additional criterion was applied in order for constructed-response items to be included as equating or anchor items. For each potential equating item, a sample of 200 papers from the 2000-01 test was randomly selected and rescored by this year's scorers. The scores for the two years were compared, and items for which there was a large difference between the average scores were excluded as equating items.

Figure 12-1
Sample Delta Plot
(♦ MC ▲ SA ● OR)



ITEM CALIBRATIONS

Common and matrix items from MEA 2000-01 were calibrated using IRT. Typically, the two-parameter logistic (2PL) model was used for dichotomous items, along with the graded response model (GRM) for the constructed-response items. Each of these models expresses the likelihood that an examinee will achieve a certain score on a set of items measuring a particular trait as a function of a parameter that is not directly observed. This parameter is commonly referred to as θ and represents

a given student's ability on the trait being measured. Using Parscale, Version 3.2, item parameters were estimated based on those models.

To calibrate items for 2001-02, parameters for the set of equating items were fixed to their calibrated values as calculated above for the 2000-01 test. This ensures that the tests for the two years are calibrated to the same ability scale. The item parameters resulting from the calibration become the basis for equated scores.

Items for 2001-02 writing were calibrated using the same method described above, except that the "equating test" consisted of the reading test, rather than a set of common writing items. Items on the 2000-01 "test" (i.e., the set of reading and writing items) were calibrated as described above. The parameters for the reading test (which was used as the equating test) were then fixed to their 2000-01 calibrated values and the 2001-02 writing items were calibrated to that same scale.

SCALED SCORES FOR READING, MATHEMATICS, SCIENCE AND TECHNOLOGY, AND SOCIAL STUDIES

For reading, mathematics, science and technology, and social studies, IRT parameters resulting from the calibrations were used to estimate student abilities. The estimated student abilities are based only on common items. The cumulative distributions of raw scores and estimated ability scores for each subject and grade combination for 2001-02 and 2000-01 were used to find the equated cutpoints. Thus, for the 2001-02 MEA a new set of cutpoints was obtained. This process is described using Figure 12-2.

Suppose $c_{2000\text{-}01}$ is a cutpoint established in 2000-01. This cutpoint is in the raw score metric. Using the frequency distribution of the raw scores for 2000-01, the cumulative percentage associated with this cutpoint was estimated through linear interpolation. Using the frequency distribution of ability estimates, the θ value associated with this cumulative percentage was determined. Because ability for 2000-01 and 2001-02 are on the same θ scale, the obtained θ value corresponds to the

same ability for both years. The 2001-02 cumulative percentage associated with this θ was then mapped to a 2001-02 raw score through linear interpolation resulting in $c_{2001\text{-}02}$.

The above process was used for each cutpoint set in 2000-01 for each grade for reading, mathematics, science an technology, and social studies. The resulting cutpoints for 2001-02 are presented in Table 12-1. These cutpoints were used to obtain new scaling parameters m_1, m_2, b_1, and b_2 which are then used to compute the scaled scores for 2001-02. The new scaling parameters are presented in Table 12-2.

The functions that translate raw scores to scaled scores are:

$S = m_1 r + b_1$	if $r < P$, and
$S = m_2 r + b_2$	if $r > P$

where S is the scaled score, r is the raw score, and P is the threshold for "Meets the Standard."

SCALED SCORES FOR WRITING

Using reading as the anchor test, 2001-02 writing raw scores were equated to 2000-01 writing raw scores using the method described above for reading, mathematics, science and technology, and social studies. However, instead of using the cumulative distributions to determine the new cutpoints as shown in Figure 12-2, the test characteristic curves (TCCs) were used. A TCC shows the relationship between student ability, θ, and expected scores on a particular test. Because ability for the two years is on the same θ scale, the new cutpoints can be determined directly from the two TCCs. This process is illustrated for the Grade 4 "meets the standard" cutpoint in Figure 12-3. The cutpoint for meeting the standard established in 2000-01 was 20.32. First, we drew a line from the 2000-01 Expected Score of 20.32 (shown on the left-hand axis of the graph). That line intersects the 2000-01 TCC at a θ value of approximately 1.3. We then drew the corresponding line from the point on the 2001-02 TCC at which θ = 1.3 to the right-hand axis of the graph, yielding a 2001-02 proficient cutpoint of 18.69. This same process was then used to find the other two cutpoints for

grade 4, as well as all cutpoints for grades 8 and 11. The 2001-02 writing cutpoints are shown in Table 12.1. Once the cutpoints had been determined, they were then used to obtain the new scaling parameters, m_1, m_2, b_1, and b_2, which were then used to compute the scaled scores for 2001-02. The new scaling parameters are presented in Table 12-2.

Figure 12-2
Finding Equated Cutpoints for Reading, Mathematics, Science and Technology, and Social Studies

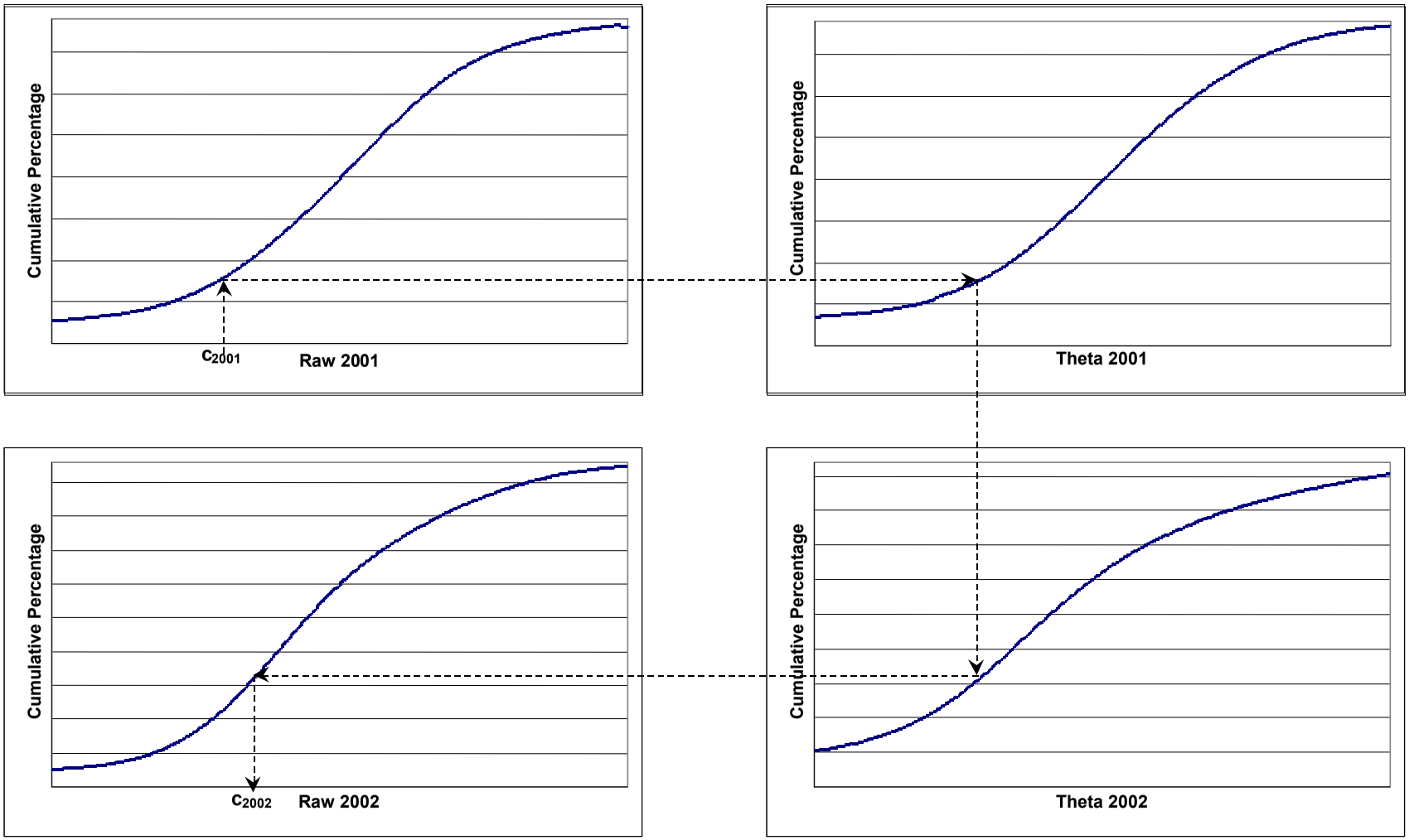


Figure 12-3
Finding Equated Cutpoints for Grade 4 Writing

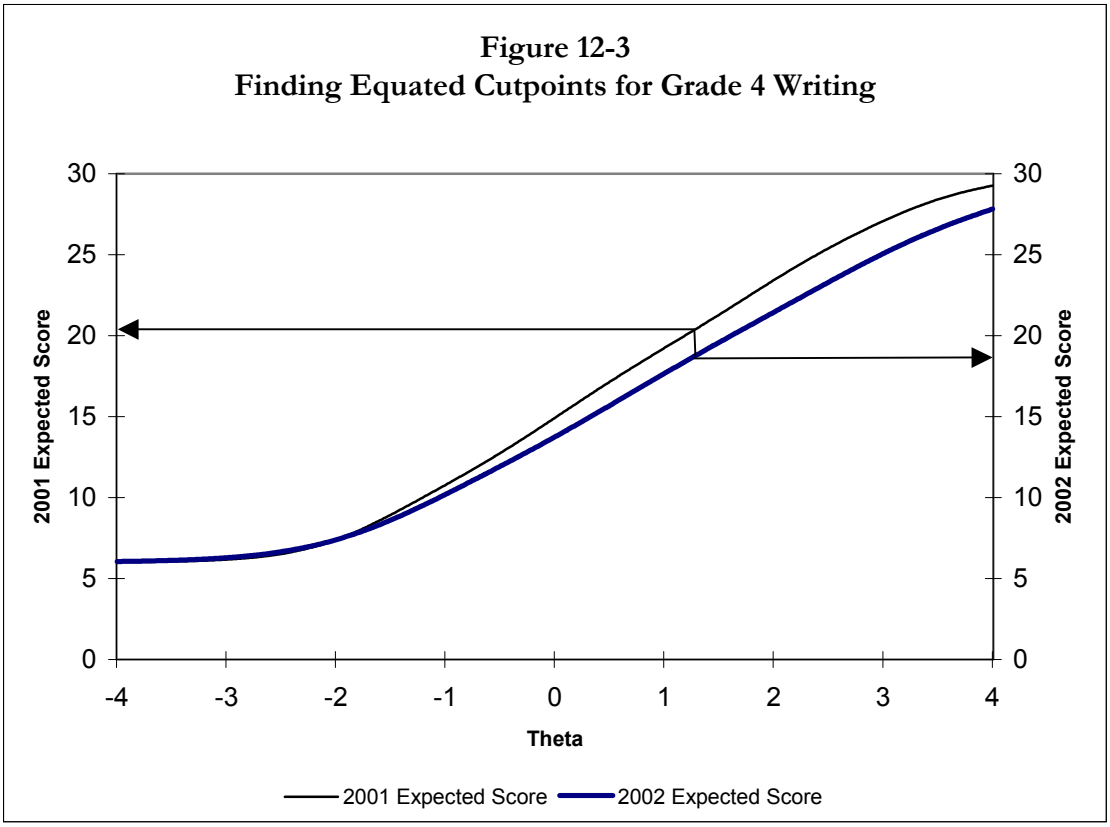


Table 12-1
Threshold (Minimum) Total Test Score For Each Performance Category for Reading, Mathematics, Science and Technology, Social Studies and Writing

Grade	Subject Area	Maximum Score on Test	Threshold Score		
			Exceeds The Standards	Meets The Standards	Partially Meets the Standards
4	Reading	52	46.60	32.88	19.77
	Mathematics	46	41.78	31.69	19.82
	Science and Technology	50	44.01	37.12	20.31
	Social Studies	50	40.98	30.83	19.16
	Writing	30	27.90	18.69	10.47
8	Reading	52	44.87	33.91	23.29
	Mathematics	46	41.47	29.75	18.08
	Science and Technology	50	40.66	31.02	18.44
	Social Studies	50	37.64	27.11	15.14
	Writing	30	27.57	18.71	11.83
11	Reading	52	45.96	34.79	23.21
	Mathematics	46	39.10	23.69	11.40
	Science and Technology	50	41.46	31.83	16.36
	Social Studies	50	38.12	25.28	15.11
	Writing	30	25.21	18.57	11.34

Table 12-2
Transformation Constants Used to Compute Scaled Scores for Reading, Mathematics, Science and Technology, Social Studies and Writing

Grade	Subject Area	Transformation Constants			
		m_1	b_1	m_2	b_2
4	Reading	1.53	490.83	1.46	493.06
	Mathematics	1.68	487.60	1.98	478.16
	Science and Technology	1.19	496.84	2.90	433.34
	Social Studies	1.71	488.17	1.97	480.24
	Writing	2.43	495.52	2.17	500.40
8	Reading	1.88	477.18	1.82	479.11
	Mathematics	1.71	490.01	1.71	490.22
	Science and Technology	1.59	491.68	2.07	476.65
	Social Studies	1.67	495.69	1.90	489.51
	Writing	2.91	486.58	2.26	498.79
11	Reading	1.73	480.90	1.79	478.76
	Mathematics	1.63	502.44	1.30	510.26
	Science and Technology	1.29	499.85	2.08	474.89
	Social Studies	1.97	491.30	1.56	501.63
	Writing	2.76	489.67	3.01	485.01

Tables 12-3 through 12-5 show the scaled score distributions for Reading, Writing, Mathematics, Science and Technology, and Social Studies.

Table 12-3
Scaled Score Distributions - Grade 4

	Reading		Writing		Mathematics		Science/Tech		Social Studies	
Score	N	%	N	%	N	%	N	%	N	%
502	78	0.51	4	0.03	524	3.34	24	0.15	78	0.49
504	79	0.51			203	1.29	11	0.07	30	0.19
506	48	0.31			212	1.35	55	0.35	70	0.44
508	70	0.45			302	1.92	133	0.84	214	1.36
510	173	1.12	532	3.46	346	2.20	119	0.75	136	0.86
512	112	0.73	473	3.08	799	5.09	420	2.66	201	1.27
514	144	0.93	732	4.76	454	2.89	663	4.20	232	1.47
516	315	2.04			517	3.29	470	2.98	289	1.83
518	219	1.42	759	4.93	524	3.34	1105	6.99	313	1.98
520	224	1.45	1054	6.85	603	3.84	1415	8.96	922	5.84
522	545	3.53	1206	7.84	1242	7.91	767	4.86	570	3.61
524	319	2.07	1822	11.85	632	4.03	1709	10.82	647	4.10
526	362	2.35			723	4.61	1806	11.43	733	4.65
528	861	5.58	1471	9.56	673	4.29	1722	10.90	747	4.74
530	526	3.41	1403	9.12	702	4.47	745	4.72	836	5.30
532	538	3.49	1201	7.81	672	4.28	1368	8.66	1716	10.88
534	595	3.86	980	6.37	1294	8.24	1128	7.14	920	5.83
536	1268	8.22	881	5.73	649	4.13	455	2.88	913	5.79
538	714	4.63			579	3.69	779	4.93	895	5.67
540	731	4.74	717	4.66	548	3.49	459	2.91	806	5.11
542	1523	9.88	601	3.91	498	3.17			767	4.86
544	778	5.04	487	3.17	510	3.25	150	0.95	747	4.74
546	783	5.08	368	2.39	418	2.66	103	0.65	665	4.22
548	1404	9.10	244	1.59	393	2.50			562	3.56
550	628	4.07	161	1.05	340	2.17	73	0.46	431	2.73
552	1072	6.95	121	0.79	306	1.95	53	0.34	371	2.35
554	382	2.48	81	0.53	266	1.69			284	1.80
556	315	2.04	36	0.23	219	1.39	30	0.19	236	1.50
558	378	2.45			178	1.13	14	0.09	175	1.11
560	108	0.70	25	0.16	136	0.87	14	0.09	109	0.69
562	62	0.40	16	0.10	94	0.60			66	0.42
564	60	0.39	5	0.03	76	0.48	4	0.03	72	0.46
566	5	0.03	1	0.01	42	0.27	1	0.01	11	0.07
568	3	0.02			18	0.11			6	0.04
570					7	0.04	1	0.01	2	0.01
572							1	0.01	3	0.02
574									1	0.01
576										
578										
580										

Note: Scaled scores that correspond to the shaded cells were unassigned.

Table 12-4
Scaled Score Distributions - Grade 8

	Reading		Writing		Mathematics		Science/Tech		Social Studies	
Score	N	%	N	%	N	%	N	%	N	%
502	172	1.03	1	0.01	945	5.63	199	1.18	120	0.71
504	63	0.38	173	1.04	334	1.99	118	0.70	91	0.54
506	70	0.42	134	0.81	397	2.37	173	1.03	174	1.03
508	118	0.70			887	5.29	233	1.38	221	1.31
510	118	0.70	231	1.39	512	3.05	612	3.63	598	3.55
512	357	2.13	233	1.40	541	3.23	406	2.41	410	2.43
514	200	1.19			551	3.29	455	2.70	463	2.75
516	312	1.86	504	3.04	568	3.39	613	3.63	573	3.40
518	339	2.02	538	3.24	630	3.76	1347	7.99	632	3.75
520	356	2.12			1315	7.84	773	4.58	1406	8.35
522	439	2.62	1014	6.11	613	3.65	854	5.06	769	4.56
524	515	3.07	928	5.59	669	3.99	881	5.22	777	4.61
526	597	3.56			626	3.73	1724	10.22	792	4.70
528	665	3.96	1063	6.40	599	3.57	853	5.06	856	5.08
530	660	3.93	1157	6.97	595	3.55	890	5.28	1634	9.70
532	782	4.66			1240	7.39	850	5.04	783	4.65
534	866	5.16	1316	7.93	570	3.40	1506	8.93	734	4.36
536	898	5.35	1285	7.74	542	3.23	678	4.02	716	4.25
538	993	5.92	1457	8.78	526	3.14	675	4.00	709	4.21
540	970	5.78			535	3.19	1119	6.64	1182	7.02
542	1885	11.24	1329	8.01	469	2.80			502	2.98
544	925	5.51	1242	7.48	812	4.84	445	2.64	497	2.95
546	906	5.40	1057	6.37	359	2.14	360	2.13	411	2.44
548	801	4.78	822	4.95	346	2.06	299	1.77	354	2.10
550	679	4.05	719	4.33	316	1.88	247	1.46	336	1.99
552	584	3.48	488	2.94	268	1.60	183	1.09	282	1.67
554	474	2.83			244	1.45	122	0.72	210	1.25
556	324	1.93	384	2.31	370	2.21	107	0.63	154	0.91
558	277	1.65	245	1.48	140	0.83	63	0.37	136	0.81
560	178	1.06	151	0.91	97	0.58	41	0.24	116	0.69
562	121	0.72	81	0.49	69	0.41	21	0.12	82	0.49
564	104	0.62	35	0.21	52	0.31	6	0.04	40	0.24
566	19	0.11	12	0.07	24	0.14	7	0.04	27	0.16
568	4	0.02			12	0.07	3	0.02	25	0.15
570	3	0.02							12	0.07
572							1	0.01	12	0.07
574									5	0.03
576									5	0.03
578									2	0.01
580										

Note: Scaled scores that correspond to the shaded cells were unassigned.

Table 12-5
Scaled Score Distributions - Grade 11

	Reading		Writing		Mathematics		Science/Tech		Social Studies	
Score	N	%	N	%	N	%	N	%	N	%
502	100	0.68	7	0.05	6	0.04	14	0.10	234	1.62
504	33	0.22			29	0.20	15	0.10	145	1.00
506	68	0.46	113	0.79	103	0.72	106	0.73	224	1.55
508	61	0.41			503	3.50	273	1.88	634	4.38
510	68	0.46	119	0.84	471	3.28	191	1.31	399	2.75
512	80	0.54	326	2.29	589	4.10	588	4.04	468	3.23
514	100	0.68	299	2.10	675	4.70	355	2.44	505	3.49
516	149	1.01			761	5.30	993	6.83	508	3.51
518	303	2.06	621	4.36	1554	10.82	613	4.21	593	4.09
520	199	1.35	591	4.15	829	5.77	1299	8.93	597	4.12
522	236	1.60	979	6.87	746	5.19	692	4.76	687	4.74
524	286	1.94			701	4.88	1474	10.13	648	4.47
526	303	2.06	842	5.91	1294	9.01	741	5.09	655	4.52
528	388	2.64	972	6.82	553	3.85	1389	9.55	657	4.54
530	426	2.89			492	3.43	1292	8.88	717	4.95
532	1096	7.45	954	6.69	527	3.67	581	3.99	659	4.55
534	606	4.12	1076	7.55	452	3.15	1040	7.15	693	4.78
536	671	4.56	999	7.01	744	5.18	471	3.24	654	4.52
538	837	5.69			328	2.28	790	5.43	664	4.58
540	909	6.18	1081	7.59	337	2.35	314	2.16	617	4.26
542	919	6.24	1010	7.09	581	4.05	276	1.90	522	3.60
544	1805	12.26			243	1.69	228	1.57	505	3.49
546	877	5.96	985	6.91	453	3.15	187	1.29	776	5.36
548	831	5.65	830	5.82	184	1.28	149	1.02	337	2.33
550	688	4.67			326	2.27	135	0.93	312	2.15
552	717	4.87	623	4.37	146	1.02	92	0.63	244	1.68
554	557	3.78	521	3.66	252	1.75	77	0.53	352	2.43
556	489	3.32			185	1.29	60	0.41	128	0.88
558	347	2.36	426	2.99	55	0.38	41	0.28	106	0.73
560	248	1.68	336	2.36	122	0.85	27	0.19	135	0.93
562	261	1.77			38	0.26	21	0.14	37	0.26
564	40	0.27	237	1.66	49	0.34	14	0.10	30	0.21
566	16	0.11	138	0.97	17	0.12	1	0.01	17	0.12
568	3	0.02			15	0.10	6	0.04	19	0.13
570	2	0.01	91	0.64			1	0.01	3	0.02
572			53	0.37			1	0.01	3	0.02
574							1	0.01	1	0.01
576			22	0.15						
578										
580										

Note: Scaled scores that correspond to the shaded cells were unassigned.

SCALED SCORES FOR HEALTH EDUCATION AND VISUAL AND PERFORMING ARTS

The equating procedure for health education and visual and performing arts is the same as that for reading, mathematics, science and technology, and social studies. However, the scaled scores for health education and visual and performing arts are linear transformations of estimated θ scores and not raw scores like in reading, mathematics, science and technology, and social studies.

The functions that translate $\hat{\theta}$s to scaled scores are

$$\begin{array}{ll} S = m_1 \hat{\theta} + b_1 & \text{if } \hat{\theta} < P, \text{ and} \\ S = m_2 \hat{\theta} + b_2 & \text{if } \hat{\theta} > P \end{array}$$

where S is the scaled score, $\hat{\theta}$ is the ability estimate found using the expected a posterior method (with a prior distribution having a mean of 0.0 and a standard deviation of 1.0), and P is the threshold for "Meets the Standard." The scaling parameters m_1, m_2, b_1, and b_2 are based on the results of standard setting processes implemented for Health Education and Visual and Performing Arts in 1999. These constants are presented in Table 12-6.

Table 12-6
Transformation Constants Used to Compute Scaled Scores for Health and Visual and Performing Arts

Grade	Subject Area	Transformation Constants			
		m_1	b_1	m_2	b_2
4	Health Education	19.68	533.95	10.13	537.37
	Visual and Performing Arts	8.21	534.14	11.40	531.48
8	Health Education	12.29	537.45	10.74	537.89
	Visual and Performing Arts	9.39	534.99	14.29	531.86
11	Health Education	13.89	536.26	10.78	537.32
	Visual and Performing Arts	5.12	536.29	14.81	527.37

CONTENT AREA SUBCATEGORY SCORES

In addition to content area scaled scores, scores for Content Area Subcategories are also provided on student score reports. These subscores are reported for reading, writing, mathematics, science and technology and social studies. Subscores are not reported for health education and visual and performing arts because individual student scores are not reported for those content areas. The subcategory scores are shown graphically on the student score reports. To compute subcategory scores, the subset of students who received a score of 542 (the lowest scaled score at which a student has met the standard) was first identified and their average score on the items comprising each subcategory was calculated. Second, the standard deviation of the subcategory scores was calculated, based on the scores of all students. Then, for each student, a standardized score (known as a z-score) could be calculated by subtracting the mean from their score and dividing that difference by the standard deviation:

$$z_X = \frac{X - \overline{X}_{542}}{s_{all}}$$

A student's z-score was positive if he/she scored above the mean, and negative otherwise.

The graph consists of a center line, which represents the mean, and three shaded bands. The innermost band marks off the area of the graph that is within one standard deviation of the mean (z from -1.0 to 1.0), the second band marks the area between one and two standard deviations from the mean (z from -1.0 to –2.0 and 1.0 to 2.0), and the third is between two and three standard deviations from the mean (z from –2.0 to –3.0 and 2.0 to 3.0). For each subcategory, the student's score was represented by a diamond printed in the appropriate place on the graph.

CHAPTER 13: ITEM ANALYSES

As noted in Brown (1983), "a test is only as good as the items it contains." A complete evaluation of a test's quality must include an evaluation of each question. Both the *Standards for Educational and Psychological Testing* and the *Code of Fair Testing Practices in Education* include standards for identifying quality questions. Questions should assess only knowledge or skills that are under assessment and should avoid assessing irrelevant factors. They should also be unambiguous and free of grammatical errors, potentially insensitive content or language, and other confounding characteristics. Further, questions must not unfairly disadvantage test takers from particular racial, ethnic, or gender groups.

Both qualitative and quantitative analyses are conducted to ensure that MEA questions meet these standards. Previous sections in this report have delineated the qualitative checks on question quality. The current chapter focuses on more quantitative evaluations. The statistical evaluations are presented in three sections: 1) difficulty indices, 2) item-test correlations, and 3) subgroup differences in item performance. The results presented in this chapter are based on the statewide administrations of the MEA in December of 2001 and March of 2002.

DIFFICULTY INDICES

All multiple-choice, short-answer, and constructed-response items were evaluated in terms of difficulty and relationship to overall score according to standard classical test theory practice. Difficulty was measured by averaging the proportion of points received across all students who received the item. Multiple-choice items were scored dichotomously (correct v. incorrect), so for these items the difficulty index is simply the proportion of students who correctly answered the item. Constructed-response items allowed for scores between zero and four. By computing the difficulty index as the average proportion of points received, the indices for multiple-choice, short-answer, and

constructed-response items are placed on a similar scale; the index ranges from zero to one regardless of the item type. Although this index is traditionally described as a measure of difficulty (as it is described here), it is properly interpreted as an easiness index because larger values indicate easier items. An index of zero indicates that no student received credit for the item, and an index of one indicates that every student received full credit for the item.

Items that are answered correctly by almost all students provide little information about differences in student ability, but they do indicate knowledge or skills that have been mastered by most students. Similarly, items that are correctly answered by very few students may indicate knowledge or skills that have not yet been mastered by most students, but such items provide little information about differences in student ability. In general, to provide best measurement, difficulty indices should range from near-chance performance (.25 for four-option, multiple-choice items or essentially zero for short-answer and open-response items) to .90. Indices outside this range indicate items that were either too difficult or too easy for the target population.

Although difficulty is an important item characteristic, the relationship between performance on a item and performance on the whole test or a relevant test section may be more critical. A item that assesses relevant knowledge or skills should relate to other items that are purported to be measuring the same knowledge or skills.

ITEM-TEST CORRELATIONS

Within classical test theory, these relationships are assessed using correlation coefficients that are typically described as either item-test correlations or, more commonly, discrimination indices. The discrimination index used to analyze MEA multiple-choice items was the point-biserial correlation between item score and a criterion total score on the test. As such, the index ranges from –1 to 1, with the magnitude and sign of the index indicating the relationship's strength and direction, respectively. For constructed-response items, item discrimination indices were based on the Pearson

product-moment correlation. The theoretical range of these statistics is also from –1 to 1, with a typical range from .3 to .6.

In general, discrimination indices are interpreted as indicating the degree to which high- and low-ability students perform differently on a item or, equivalently, the degree to which performance on a item helps to differentiate between high- and low-ability students. From this perspective, indices near 1 indicate that high-ability students are more likely to answer the item correctly, indices near –1 indicate that low-ability students are more likely to answer the item correctly, and indices near 0 indicate that the item is equally likely to be answered correctly by high- and low-ability students.

Discrimination indices can be thought of as measures of how closely a item assesses the same knowledge and skills assessed by other items contributing to the criterion total score; that is, the discrimination index can be interpreted as a measure of construct consistency. In light of this interpretation, the selection of an appropriate criterion total score is crucial to the interpretation of the discrimination index. For the 2001-02 MEA the criterion score for each common item is the total score for all common items. For each matrix item the criterion score is the total score for the form that item belongs to.

SUMMARY OF ITEM ANALYSIS RESULTS

Summary statistics of the difficulty and discrimination indices for each item are provided in Tables 13-1 through13-3. In general, the item difficulty and discrimination indices are in acceptable and expected ranges. Very few items were answered correctly at near-chance or near-perfect rates. Similarly, the positive discrimination indices indicate that most items were assessing consistent constructs, and students who performed well on individual items tended to perform well overall. There was a small number of items with near-zero discrimination indices, but none was reliably negative. Occasionally, items with less-desirable statistical characteristics need to be included in

assessments to ensure that content is appropriately covered, but there were very few such cases on MEA.

A comparison of indices across grade levels is complicated because these indices are population dependent. Direct comparisons would require that either the items or students were common across groups. However, one can say that with respect to multiple-choice items the fourth- and eighth-grade students tended to have less difficulty answering the mathematics items on the fourth- and eighth-grade tests as compared to the eleventh-grade students answering the mathematics items on the eleventh-grade tests. This trend was also found across the other content areas, except for health where $p = 0.71$ for both grades 8 and 11.

Comparing the difficulty indices of multiple-choice and short-answer or constructed-response items is inappropriate because multiple-choice items can be answered correctly by guessing. Thus, it is not surprising that the difficulty indices for multiple-choice items tend to be higher (indicating easier items) than the difficulty indices for other item types. Similarly, the partial credit allowed for constructed-response items is advantageous in the computation of item-test correlations, so the discrimination indices for these items tend to be larger than the discrimination indices of other item types.

Table 13-1
Average Difficulty and Discrimination of Different Item Types For Each Grade-Content Area Combination - Grade 4

Content Area	Statistics	Item Type		
		All	Multiple Choice	Open Response
Reading	Difficulty	0.65 (0.18)	0.71 (0.15)	0.46 (0.12)
	Discrimination	0.39 (0.11)	0.35 (0.09)	0.50 (0.07)
	N	133	104	29
Mathematics	Difficulty	0.58 (0.17)	0.63 (0.15)	0.43 (0.14)
	Discrimination	0.37 (0.10)	0.33 (0.08)	0.48 (0.07)
	N	125	92	33
Science and Technology	Difficulty	0.63 (0.19)	0.68 (0.17)	0.40 (0.16)
	Discrimination	0.27 (0.09)	0.25 (0.08)	0.36 (0.11)
	N	138	116	22
Social Studies	Difficulty	0.58 (0.19)	0.61 (0.18)	0.42 (0.21)
	Discrimination	0.29 (0.10)	0.27 (0.09)	0.37 (0.10)
	N	138	116	22
Health	Difficulty	0.62 (0.16)	0.66 (0.16)	0.48 (0.08)
	Discrimination	0.21 (0.06)	0.19 (0.06)	0.25 (0.05)
	N	128	96	32
VPA	Difficulty	0.58 (0.15)	0.61 (0.14)	0.43 (0.07)
	Discrimination	0.22 (0.06)	0.21 (0.06)	0.24 (0.05)
	N	84	72	12

Table 13-2
Average Difficulty and Discrimination of Different Item Types For Each Grade-Content Area Combination - Grade 8

Content Area	Statistics	Item Type		
		All	Multiple Choice	Open Response
Reading	Difficulty	0.66 (0.16)	0.71 (0.14)	0.50 (0.08)
	Discrimination	0.36 (0.12)	0.32 (0.09)	0.53 (0.09)
	N	133	104	29
Mathematics	Difficulty	0.47 (0.16)	0.51 (0.15)	0.36 (0.15)
	Discrimination	0.40(0.12)	0.36 (0.09)	0.53 (0.08)
	N	125	92	33
Science and Technology	Difficulty	0.57 (0.21)	0.62 (0.19)	0.32 (0.13)
	Discrimination	0.30 (0.11)	0.27 (0.09)	0.45 (0.06)
	N	138	116	22
Social Studies	Difficulty	0.56 (0.18)	0.60 (0.16)	0.35 (0.14)
	Discrimination	0.34 (0.11)	0.31 (0.09)	0.49 (0.09)
	N	138	116	22
Health	Difficulty	0.64 (0.18)	0.71 (0.15)	0.45 (0.08)
	Discrimination	0.25 (0.09)	0.22 (0.08)	0.35 (0.07)
	N	128	96	32
VPA	Difficulty	0.59 (0.17)	0.62 (0.16)	0.40 (0.04)
	Discrimination	0.26 (0.07)	0.25 (0.07)	0.31 (0.06)
	N	84	72	12

Table 13-3 Average Difficulty and Discrimination of Different Item Types For Each Grade-Content Area Combination – Grade 11				
		Item Type		
Content Area	Statistics	All	Multiple Choice	Open Response
Reading	Difficulty	0.66 (0.17)	0.69 (0.17)	0.56 (0.13)
	Discrimination	0.35 (0.15)	0.30 (0.12)	0.54 (0.10)
	N	133	104	29
Mathematics	Difficulty	0.41 (0.17)	0.45 (0.16)	0.29 (0.13)
	Discrimination	0.39 (0.14)	0.33 (0.09)	0.56 (0.11)
	N	125	92	33
Science and Technology	Difficulty	0.50 (0.19)	0.53 (0.19)	0.34 (0.08)
	Discrimination	0.31 (0.12)	0.28 (0.09)	0.50 (0.08)
	N	138	116	22
Social Studies	Difficulty	0.55 (0.18)	0.59 (0.16)	0.37 (0.11)
	Discrimination	0.34 (0.13)	0.30 (0.10)	0.53 (0.06)
	N	138	116	22
Health	Difficulty	0.66 (0.15)	0.71 (0.14)	0.51 (0.06)
	Discrimination	0.23 (0.10)	0.18 (0.06)	0.37 (0.07)
	N	128	96	32
VPA	Difficulty	0.58 (0.17)	0.60 (0.17)	0.45 (0.08)
	Discrimination	0.27 (0.07)	0.25 (0.07)	0.36 (0.04)
	N	84	72	12

SUBGROUP DIFFERENCES IN ITEM PERFORMANCE

The Code of Fair Testing Practices in Education explicitly states that subgroup differences in performance should be examined when sample sizes permit, and actions should be taken to make certain that differences in performance are due to construct-relevant, rather than irrelevant, factors. The *Standards for Educational and Psychological Testing* includes similar guidelines. As part of the effort to identify such problems, MEA items were evaluated in terms of differential item functioning (DIF) statistics.

DIF procedures are designed to identify items for which subgroups of interest perform differently beyond the impact of differences in overall achievement. For the MEA, the standardization DIF procedure (Dorans and Kulick, 1986) was employed to evaluate subgroup differences between male and female. This procedure calculates the difference in item performance

for groups of students matched for achievement on the total test. That is, the average item performance is calculated for students at every total score, then an overall average is calculated weighting the total score distribution so it is the same for the two groups.

The index ranges from –1 to 1 for multiple-choice and short-answer items and is adjusted to the same scale for constructed-response items. Negative numbers indicate that the item was more difficult for females. Positive numbers indicate that the item was easier for female students. Dorans and Holland (1993) suggested that index values between –0.05 and 0.05 should be considered negligible for dichotomously scored items (such as MEA multiple-choice items). Most MEA items fall within this range. Dorans and Holland further stated that dichotomously scored items with values between –0.10 and –0.05 and between 0.05 and 0.10 (i.e., "low" DIF) should be inspected to ensure that no possible effect is overlooked, and that items with values outside the [–0.10, 0.10] range (i.e., "high" DIF) are more unusual and should be examined very carefully. These standards can be applied to constructed-response items by accounting for the larger range of possible index values and scaling appropriately. That is, values of the DIF index for open-response items can range from –4.0 to 4.0, so the corresponding ranges are between -0.2 and 0.2 for negligible difference, between –0.4 and –0.2 and between 0.2 and 0.4 for "low" DIF, and outside [-0.4, 0.4] for "high" DIF.

DIF indices indicate differential performance between two groups. That differential performance may or may not be indicative of bias in the test. Course-taking patterns, group differences in interests, or differences in school curricula can lead to DIF. If subgroup differences in performance are related to construct-relevant factors, the items should be considered for inclusion on a test.

Each item was categorized according to the guidelines adapted from Dorans and Holland (1993). Tables 13-4 to 13-6 provide the number of items in each of the three DIF categories for male v. female for each grade level tested. There are some MEA items categorized as "low" or "high" DIF. These indices must not be interpreted as indisputable evidence of bias. Both the *Code of Fair*

Testing Practices in Education and the *Standards for Educational and Psychological Testing* assert that test items must be free from construct-irrelevant sources of differential difficulty. If subgroup differences in performance can be plausibly attributed to construct-relevant factors, the items may be included on a test. What is important is to determine if the cause of this differential performance is construct relevant.

Table 13-4
Differential Item Functioning (DIF) Categorization Item Type: Grade 4

Content Area	Item Type	Negligible DIF				Low DIF				High DIF			
		Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%
Reading	Multiple Choice	41	44	85	0.82	2	15	17	0.16	0	2	2	0.02
	Constructed Response	26	3	29	1.00	0	0	0	0.00	0	0	0	0.00
Mathematics	Multiple Choice	47	29	76	0.83	4	11	15	0.16	1	0	1	0.01
	Constructed Response	18	11	29	0.88	2	2	4	0.12	0	0	0	0.00
Science and Technology	Multiple Choice	51	34	85	0.73	8	18	26	0.22	1	4	5	0.04
	Constructed Response	14	6	20	0.91	0	2	2	0.09	0	0	0	0.00
Social Studies	Multiple Choice	40	57	97	0.84	3	13	16	0.14	0	3	3	0.03
	Constructed Response	16	5	21	0.95	1	0	1	0.05	0	0	0	0.00
Health	Multiple Choice	37	35	72	0.75	7	13	20	0.21	2	2	4	0.04
	Constructed Response	21	2	23	0.72	8	0	8	0.25	1	0	1	0.03
VPA	Multiple Choice	32	25	57	0.79	6	5	11	0.15	2	2	4	0.06
	Constructed Response	4	0	4	0.33	7	0	7	0.58	1	0	1	0.08

Table 13-5
Differential Item Functioning (DIF) Categorization Item Type: Grade 8

Content Area	Item Type	Negligible DIF				Low DIF				High DIF			
		Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%
Reading	Multiple Choice	34	47	81	0.78	4	17	21	0.20	0	2	2	0.02
	Constructed Response	21	2	23	0.79	6	0	6	0.21	0	0	0	0.00
Mathematics	Multiple Choice	27	40	67	0.73	3	19	22	0.24	0	3	3	0.03
	Constructed Response	21	8	29	0.88	1	3	4	0.12	0	0	0	0.00
Science and Technology	Multiple Choice	42	45	87	0.75	5	20	25	0.22	1	3	4	0.03
	Constructed Response	12	5	17	0.77	4	1	5	0.23	0	0	0	0.00
Social Studies	Multiple Choice	40	48	88	0.76	6	20	26	0.22	0	2	2	0.02
	Constructed Response	10	5	15	0.68	4	0	4	0.18	3	0	3	0.14
Health	Multiple Choice	37	31	68	0.71	8	16	24	0.25	0	4	4	0.04
	Constructed Response	14	1	15	0.47	16	0	16	0.50	1	0	1	0.03
VPA	Multiple Choice	28	21	49	0.68	7	10	17	0.24	4	2	6	0.08
	Constructed Response	1	0	1	0.08	8	0	8	0.67	3	0	3	0.25

Table 13-6
Differential Item Functioning (DIF) Categorization Item Type: Grade 11

Content Area	Item Type	Negligible DIF				Low DIF				High DIF			
		Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%	Favor Female	Favor Male	N	%
Reading	Multiple Choice	33	46	79	0.76	2	22	24	0.23	0	1	1	0.01
	Constructed Response	17	1	18	0.62	11	0	11	0.38	0	0	0	0.00
Mathematics	Multiple Choice	41	32	73	0.79	4	12	16	0.17	0	3	3	0.03
	Constructed Response	20	7	27	0.82	3	3	6	0.18	0	0	0	0.00
Science and Technology	Multiple Choice	44	30	74	0.64	6	25	31	0.27	0	11	11	0.09
	Constructed Response	9	3	12	0.55	6	3	9	0.41	1	0	1	0.05
Social Studies	Multiple Choice	43	43	86	0.74	12	12	24	0.21	0	6	6	0.05
	Constructed Response	9	3	12	0.55	7	1	8	0.36	1	1	2	0.09
Health	Multiple Choice	25	24	49	0.51	8	11	19	0.20	4	24	28	0.29
	Constructed Response	9	0	9	0.28	15	0	15	0.47	2	6	8	0.25
VPA	Multiple Choice	31	19	50	0.69	9	6	15	0.21	2	5	7	0.10
	Constructed Response	0	0	0	0.00	5	0	5	0.42	7	0	7	0.58

CHAPTER 14: RELIABILITY

Although an individual item's performance is an important focus for evaluation, a complete evaluation of an assessment must also address the way that items function together and complement one another. Any measurement includes some amount of measurement error; that is, no measurement can be perfectly accurate. This is true of academic assessments—no assessment can measure students with perfect accuracy; some students will receive scores that underestimate their true ability, and other students will receive scores that overestimate their true ability. Items that function well together produce assessments that have less measurement error; that is, the errors made should be small on average. Such assessments are described as reliable.

There are a number of ways to estimate an assessment's reliability. One approach is to split all test items into two groups and then correlate students' scores on the two half tests. This is known as a split-half estimate of reliability. If the two half-test scores correlate highly, items on the two half tests must be measuring very similar knowledge or skills. This is evidence that the items complement one another and function well as a group. This also suggests that measurement error will be minimal.

The split-half method requires the psychometrician to select which items contribute to each half-test score. This decision may have an impact on the resulting correlation. Cronbach (1951) provided a statistic that avoids this concern about the split-half method. Cronbach's α coefficient is an estimate of the average of all possible split-half reliability coefficients.

RELIABILITY AND STANDARD ERRORS OF MEASUREMENT

Table 14-1 presents descriptive statistics, Cronbach's α coefficient, and raw and scaled score standard errors of measurement for each subject separately for each grade level. Cronbach's α is computed using the following formula:

$$\alpha \equiv \frac{n}{n-1}\left[1-\frac{\sum_{i=1}^{n}\sigma^2\left(Y_i\right)}{\sigma_x^2}\right]$$

where i indexes the item
n is the total number of items,
$\sigma^2\left(Y_i\right)$ represents individual item variance, and
σ_x^2 represents the total test variance

The reported reliabilities for health and VPA are the averages of the computed Cronbach's α across forms. Because it is inappropriate to compute averages of correlations directly, Fisher's Z transformation was used. The average of the Zs was calculated, and the average was transformed back into a correlation coefficient. The low reliability values for health and VPA seen in Table 14-1 can be attributed to the lower number of items in each form in those tests.

Note that two scaled-score standard errors of measurement are presented: one for scaled scores below 542 and one for scaled scores of 542 and above. This is because different slopes were used in the linear transformation to scaled scores at these two different parts of the scaled score range.

Table 14-1
Reliabilities, Standard Errors of Measurement and Descriptive Statistics
MEA 2001-2002

Grade	Content Area	n	Points	Min	Max	Mean	S.D.	Rel.	S.E.M.	Scaled Score <542 S.E.M.	Scaled Score >=542 S.E.M.
4	Reading	15422	52	2	51	31.23	8.19	0.86	3.09	5.37	4.64
	Writing	15381	30	2	30	13.67	4.25	0.63	2.60	4.79	2.62
	Mathematics	15699	46	0	46	24.63	8.37	0.83	3.42	5.43	5.47
	Science/Tech	15797	50	3	48	24.69	6.81	0.75	3.43	4.10	8.32
	Social Studies	15776	50	2	48	26.39	6.75	0.77	3.24	5.94	6.46
	Health*	15744	14	0	14	7.82	2.21	0.45	1.65	3.75	4.92
	VPA*	15742	10	0	10	5.39	1.93	0.44	1.44	8.31	6.53
8	Reading	16774	52	4	50	31.75	6.95	0.82	2.95	6.63	5.83
	Writing	16599	30	2	30	17.16	4.59	0.60	2.90	5.50	3.33
	Mathematics	16773	46	0	46	21.57	9.16	0.88	3.22	5.26	4.93
	Science/Tech	16864	50	0	46	22.63	7.08	0.79	3.24	5.41	6.79
	Social Studies	16848	50	0	47	20.48	7.63	0.83	3.10	5.52	6.45
	Health*	16899	14	0	14	7.72	2.44	0.51	1.70	3.94	3.24
	VPA*	16742	10	0	10	5.32	2.01	0.50	1.42	8.57	6.14
11	Reading	14719	52	4	51	34.16	6.97	0.82	2.93	6.05	5.50
	Writing	14251	30	2	30	16.78	4.85	0.65	2.85	5.01	4.50
	Mathematics	14360	46	0	45	15.64	8.57	0.86	3.17	4.72	4.11
	Science/Tech	14548	50	0	48	20.83	7.57	0.82	3.25	4.33	6.73
	Social Studies	14485	50	0	46	20.05	7.65	0.82	3.25	6.33	5.38
	Health*	14802	14	0	14	8.36	2.23	0.49	1.59	4.11	3.55
	VPA*	14374	10	0	10	5.42	2.10	0.51	1.47	8.56	3.29

*The reported reliability is the average reliability across forms.

The standard error of measurement of each content area test was taken into consideration when reporting individual student scores. These standard errors were computed at each raw score level and used to report error bands around scaled scores. The standard error for a student with a raw score of *Y* was found by using the following formula (Lord & Novick, 1968):

$$se_y = \sqrt{\frac{(n - Y)(Y)}{n - 1}}$$

where n is the total possible raw score. The value of the standard error was then subtracted from and added to the raw score, giving a raw score error band. For purposes of reporting, each raw score and its upper and lower error band limits were then scaled using the appropriate trans-formation constants. (The scaling process is described in Chapter 12, and the transformation constants can be found in Table 12-2.) If either the upper or lower limit of the error band was outside the range of possible scaled scores, the confidence interval was truncated accordingly. In other words, if the upper limit of the error band for a given score was greater than the highest possible scaled score, the upper limit was set equal to that score.

STRATIFIED COEFFICIENT α

According to Feldt and Brennan (1989) a prescribed distribution of items over categories (such as different item types) indicates the presumption that at least a small, but important, degree of unique variance is associated with the categories. In contrast, Cronbach's coefficient α is built upon the assumption that there are no such local or clustered dependencies. A stratified version of coefficient α corrects for this problem:

$$\alpha_{strat} = 1 - \frac{\sum_{j=1}^{k} \sigma_{x_j}^2 (1-\alpha)}{\sigma_x^2}$$

where j indexes the subtests or categories,

$\sigma_{x_j}^2$ represents the variance of the k individual subtests or categories,

α is the unstratified Cronbach's α coefficient, and

σ_x^2 represents the total test variance

Stratified coefficient α was calculated separately for each common item test and grade level. The stratification was based on item types (multiple choice v. constructed response). These results are provided in Table 14-2.

Table 14-2 Coefficients α and Stratified α MEA 2001–2002							
Grade	Subject	α	α_{mc}	N_{mc}	α_{cr}	N_{cr}	Stratified α
4	Reading	0.86	0.81	24	0.74	9 (28)	0.87
	Mathematics	0.83	0.74	20	0.74	9 (26)	0.84
	Science/Tech	0.75	0.65	20	0.61	10 (30)	0.75
	Social Studies	0.77	0.71	20	0.61	10 (30)	0.78
8	Reading	0.82	0.72	24	0.73	9 (28)	0.83
	Mathematics	0.88	0.81	20	0.79	9 (26)	0.88
	Science/Tech	0.79	0.58	20	0.75	10 (30)	0.81
	Social Studies	0.83	0.70	20	0.77	10 (30)	0.84
11	Reading	0.82	0.70	24	0.76	9 (28)	0.83
	Mathematics	0.86	0.70	20	0.83	9 (26)	0.88
	Science/Tech	0.82	0.63	20	0.77	10 (30)	0.83
	Social Studies	0.82	0.54	20	0.82	10 (30)	0.84

RELIABILITY OF PERFORMANCE LEVEL CATEGORIZATION

All test scores contain measurement error; thus classifications based on test scores are also subject to measurement error. After the performance levels were specified and students were classified into those levels, empirical analyses were conducted to determine the statistical accuracy and consistency of the classifications.

ACCURACY

Accuracy refers to the extent to which decisions based on test scores match decisions that would have been made if the scores did not contain any measurement error. Accuracy must be estimated because errorless test scores do not exist.

CONSISTENCY

Consistency measures the extent to which classification decisions based on test scores match the decisions based on scores from a second, parallel, form of the same test. Consistency can be evaluated directly from actual responses to test items if two complete, parallel, forms of the test are given to the same group of students. This is usually impractical, especially on lengthy tests such as

the MEA. To overcome this issue, techniques have been developed to estimate both accuracy and consistency of classification decisions based on a single administration of a test. The technique developed by Livingston and Lewis (1995) was used for the MEA because their technique can be used with both constructed-response and multiple-choice items.

CALCULATING ACCURACY

All of the accuracy and consistency estimation techniques described below make use of the concept of "true scores" in the sense of classical test theory. A true score is the score that would be obtained on a test that had no measurement error. It is a theoretical concept that cannot be observed, although it can be estimated. Following Livingston and Lewis (1995), the true-score distribution for the MEA was estimated using a four-parameter beta distribution, which is a flexible model that allows for extreme degrees of skewness in test scores.

In the Livingston and Lewis method, the estimated "true scores" are used to classify students into their "true" performance category, which is labeled "true status." After various technical adjustments (which are described in Livingston and Lewis, 1995), a 4×4 contingency table was created for each content area test and grade level. The cells in the table are the proportion of students who were classified into each performance category by the actual (or observed) scores on the MEA (i.e., observed status) and by the "true scores" (i.e., "true status"). As an example, Table 14-3 shows the accuracy contingency table for fourth-grade science an technology. The accuracy contingency tables for all grades and subjects are provided in Appendix C (under step 5). Additional steps in the analysis are also shown in Appendix C.

Table 14-3 Accuracy Contingency Table for Grade 4 Science and Technology				
True Status	Observed Status			
	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards
Does Not Meet the Standards	**0.21**	0.08	0.00	0.00
Partially Meets the Standards	0.07	**0.61**	0.02	0.00
Meets the Standards	0.00	0.01	**0.01**	0.00
Exceeds the Standards	0.00	0.00	0.00	**0.00**

Proportions on the diagonal (in bold) indicate exact agreement between the observed status and "true status." If the test were perfectly accurate, all of the off-diagonal cells would be zero. Accuracy is the sum of the diagonal (i.e., the proportion of exact agreement across the four performance levels). In Table 14-3, the diagonal sums to .83, indicating that 83 percent of the students were classified into exactly the same performance categories by their observed scores and their "true scores."

CALCULATING CONSISTENCY

To estimate consistency, the "true scores" are used to estimate the distribution of classifications on an independent, parallel test form. After statistical adjustments (see Livingston and Lewis, 1995), a new 4 × 4 contingency table was created for each test and grade level that shows the proportions of students who were classified into each performance category by the actual test and by another (hypothetical) parallel test form. Consistency, which is the proportion of students classified into exactly the same categories by the two forms of the test, is the sum of the diagonal for the new contingency table. The consistency contingency tables are shown under step 7 in Appendix C.

KAPPA

Another way to measure consistency is to use Cohen's (1960) coefficient κ (kappa), which assesses the proportion of consistent classifications after removing the proportion of consistent classification that would be expected by chance. Cohen's κ can be used to estimate the classification

consistency of a test from two parallel forms of the test. The second form in this case was the one estimated using the Livingston and Lewis (1995) method. Cohen's κ is shown in Table 14-4. Because κ is corrected for chance, the values of κ are lower than the other consistency estimates in Table 14-4.

RESULTS OF ACCURACY, CONSISTENCY, AND KAPPA ANALYSES

The accuracy, consistency, and kappa indices for all grades and subjects are summarized in Table 14-4.

Table 14-4 Estimates of Accuracy and Consistency of Performance Level Classification				
Grade	Subject	Accuracy	Consistency	Kappa (κ)
4	Reading	0.79	0.72	0.52
	Writing	0.74	0.63	0.32
	Mathematics	0.77	0.67	0.48
	Science and Technology	0.83	0.75	0.47
	Social Studies	0.77	0.67	0.44
	Health	0.75	0.63	0.22
	Visual and Performing Arts	0.58	0.48	0.17
8	Reading	0.76	0.68	0.47
	Writing	0.69	0.57	0.28
	Mathematics	0.81	0.73	0.59
	Science and Technology	0.79	0.71	0.47
	Social Studies	0.79	0.70	0.51
	Health	0.75	0.66	0.29
	Visual and Performing Arts	0.56	0.45	0.19
11	Reading	0.77	0.69	0.47
	Writing	0.67	0.57	0.30
	Mathematics	0.81	0.73	0.58
	Science and Technology	0.82	0.74	0.52
	Social Studies	0.77	0.67	0.50
	Health	0.74	0.65	0.25
	Visual and Performing Arts	0.61	0.48	0.19

For certain tests, concern may be greatest regarding decisions made about a particular threshold. For example, if a college gave credit to students who achieved an Advanced Placement test score of four or five, but not one, two, or three, one might be interested in the accuracy of the dichotomous decision, below four versus four or above. Table 14-5 reports accuracy and consistency for various dichotomous categorizations on the MEA. MEA partially meets/meets cut accuracy ranges from .78 to .97, and meets/exceeds accuracy ranges from .96 to .99+. These are relatively high values compared to the 1999 Advanced Placement (AP) accuracy of decisions based on the 2-3 cut and 3-4 cut which range from .84 to .95.

Table 14-5
Accuracy and Consistency of Dichotomous Categorizations

Grade	Subject	Accuracy			Consistency		
		D/P*	P/M*	M/E*	D/P	P/M	M/E
4	Reading	0.92	0.89	0.98	0.90	0.84	0.97
	Writing	0.85	0.89	0.99+	0.77	0.85	0.99+
	Mathematics	0.89	0.89	0.99	0.84	0.85	0.98
	Science and Technology	0.86	0.97	0.99+	0.80	0.96	0.99+
	Social Studies	0.91	0.86	0.99	0.87	0.81	0.99
	Health	0.99	0.78	0.99	0.97	0.67	0.98
	Visual and Performing Arts	0.78	0.81	0.97	0.70	0.76	0.95
8	Reading	0.91	0.87	0.98	0.88	0.82	0.97
	Writing	0.89	0.82	0.99	0.83	0.75	0.97
	Mathematics	0.91	0.91	0.99	0.87	0.88	0.99
	Science and Technology	0.87	0.93	0.99+	0.81	0.89	0.99+
	Social Studies	0.89	0.91	0.99	0.84	0.87	0.99
	Health	0.98	0.78	0.99+	0.96	0.70	0.99+
	Visual and Performing Arts	0.78	0.79	0.96	0.69	0.73	0.93
11	Reading	0.92	0.87	0.98	0.92	0.80	0.97
	Writing	0.89	0.82	0.96	0.85	0.76	0.95
	Mathematics	0.89	0.93	0.99	0.84	0.90	0.99
	Science and Technology	0.87	0.95	0.99+	0.81	0.93	0.99+
	Social Studies	0.89	0.88	0.99	0.84	0.84	0.99
	Health	0.97	0.78	0.99+	0.96	0.69	0.99+
	Visual and Performing Arts	0.72	0.87	0.99+	0.65	0.83	0.96

*D/P = Does not meet/partially meets the standards
P/M = Partially meets/meets the standards
M/E = Meets/exceeds the standards

CHAPTER 15: VALIDITY

As noted in the *Standards for Educational and Psychological Testing*, validity is the most important consideration in test evaluation. Validity refers to whether specific inferences made from test scores are appropriate, meaningful, and useful. There are several types of validity-related evidence that can be used to support appropriate, meaningful, and useful inferences based on test scores.

CONTENT-RELATED EVIDENCE

As noted in the Standards, evidence of test validity begins with test development and continues throughout the entire testing process. Chapters 2 through 9 provide evidence regarding the alignment between the content of the MEA and Maine's *Learning Results*.

EXTERNAL EVIDENCE

External validity of the MEA is conveyed by the relationship of test scores and situational variables such as school transience, course-taking pattern, attitude towards subject matter, and self-image. These situational variables were all based on student questionnaire data collected during the administration of the MEA. Note that not all the questionnaire items referred to in the following subsections were asked regarding all of the subjects assessed by the MEA. Note also that no inferential statistics are included. However, because the numbers of students are large enough, differences in average scores could be shown to be statistically significant.

SCHOOL TRANSIENCE

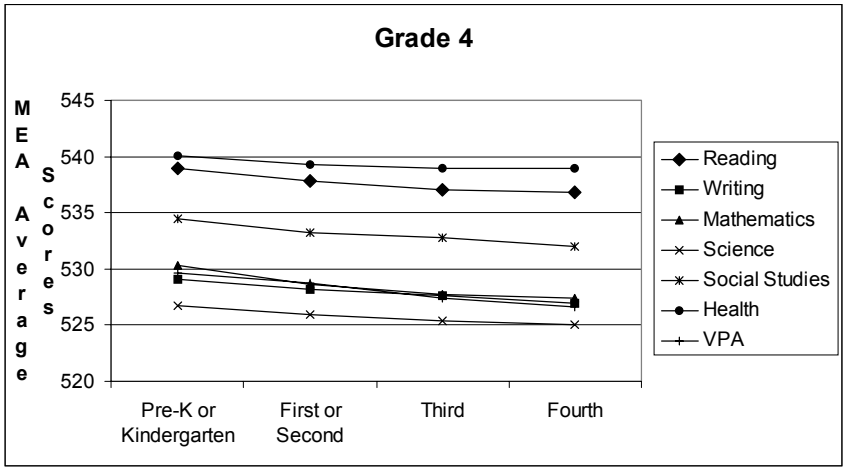
This is an evaluation of how time in a single school is related to test scores. Students were asked, "In what grade did you start coming to school in this school district?" Medsker (1998) found that typically, students who change schools often do not perform as well as students who regularly attend a single school or school system. Charts in Figure 15-1 clearly indicate that students who

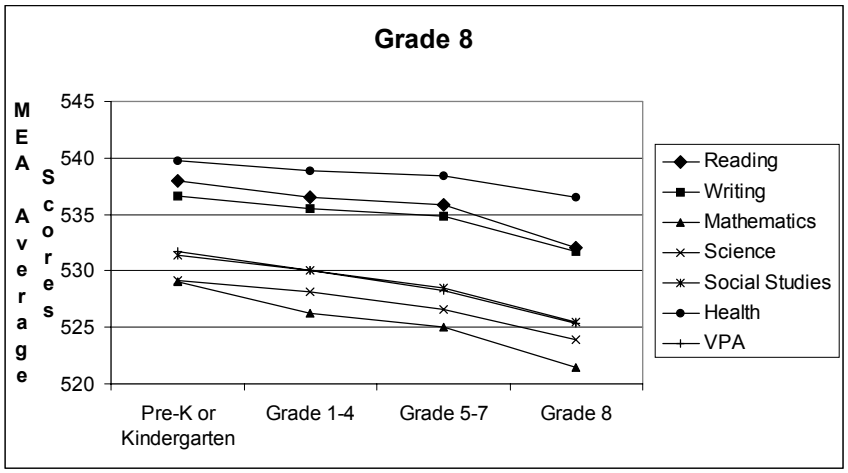
spent more time in a single school tended to have higher test scores in math, science, social studies, and visual and performing arts.

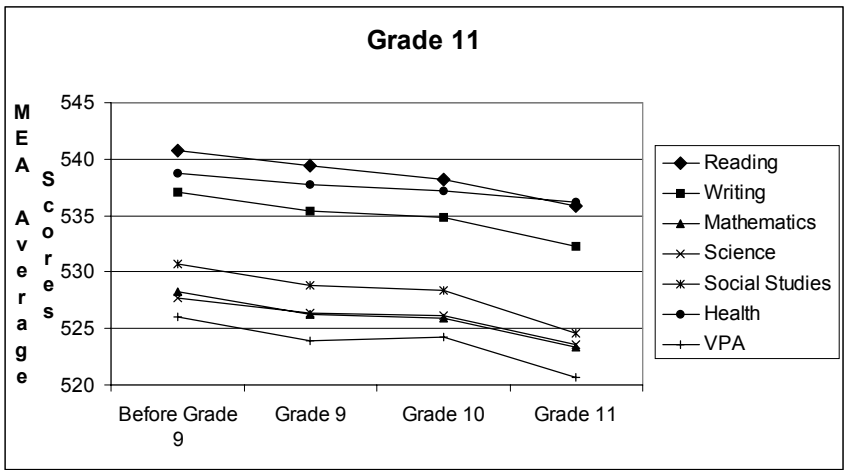
Figure 15-1

School Transience and MEA Scores

Question: In what grade did you start coming to school in this school district?







COURSE-TAKING PATTERN

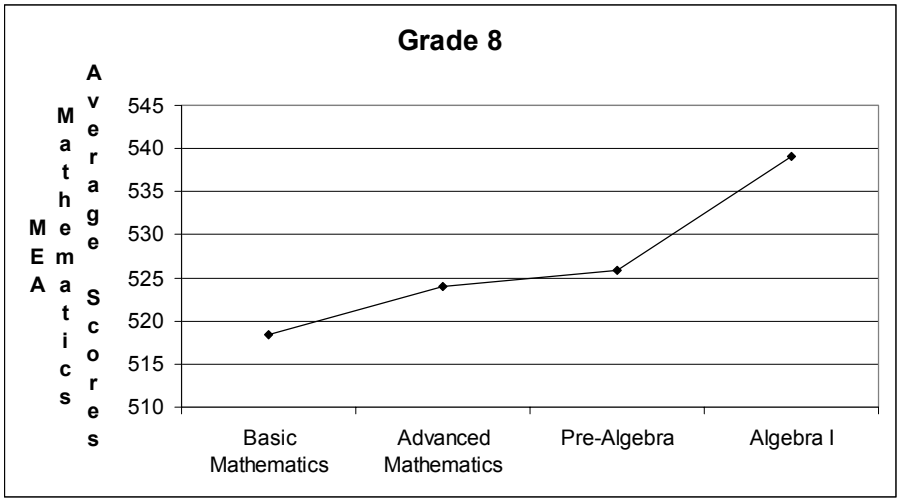
Grade 8 and 11 examinees were asked questions related to their course-taking patterns in mathematics. Eighth-graders were asked, "What best describes the mathematics class you are taking in the eighth grade?" and eleventh-graders were asked, "What mathematics courses will you complete before you graduate?" Charts in Figure 15-2 both show that the higher-level mathematics courses are associated with higher MEA mathematics scores.

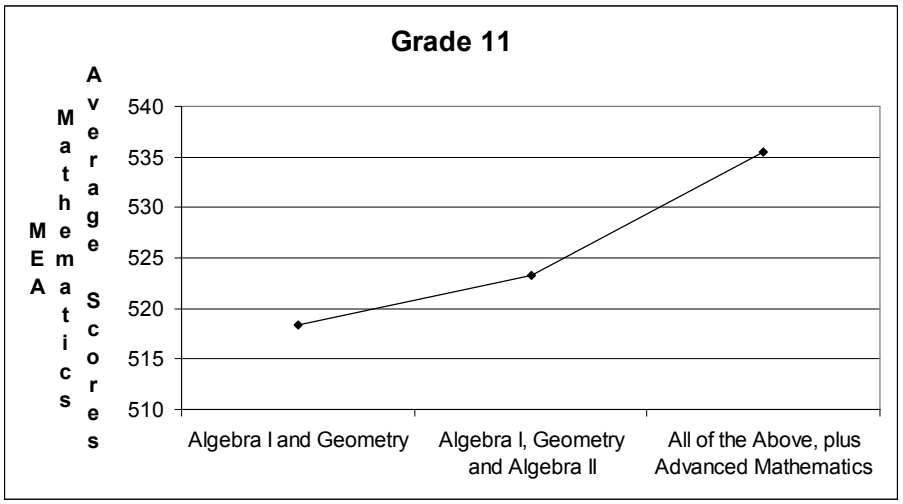
Figure 15-2

MEA Mathematics Scores and Course-Taking Patterns

Grade 8 Question: What best describes the mathematics class you are taking in the eighth grade?

Grade 11 Question: What mathematics courses will you complete before you graduate?





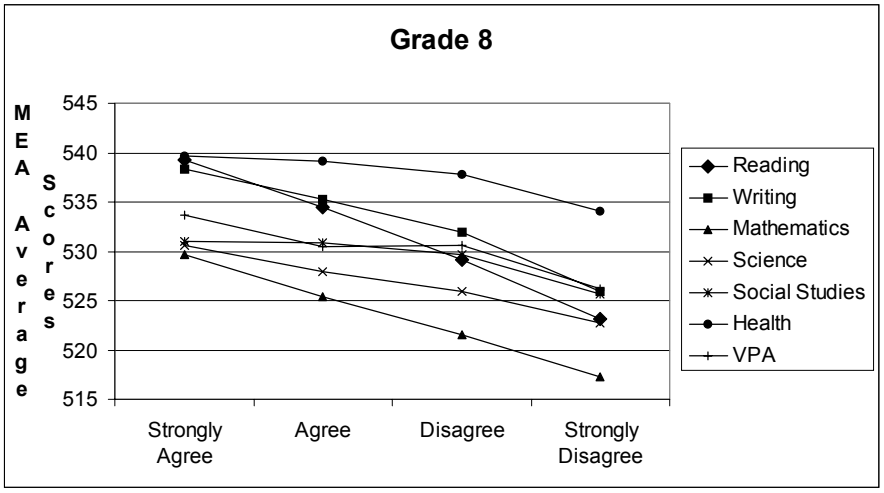
ATTITUDE TOWARDS SUBJECT MATTER

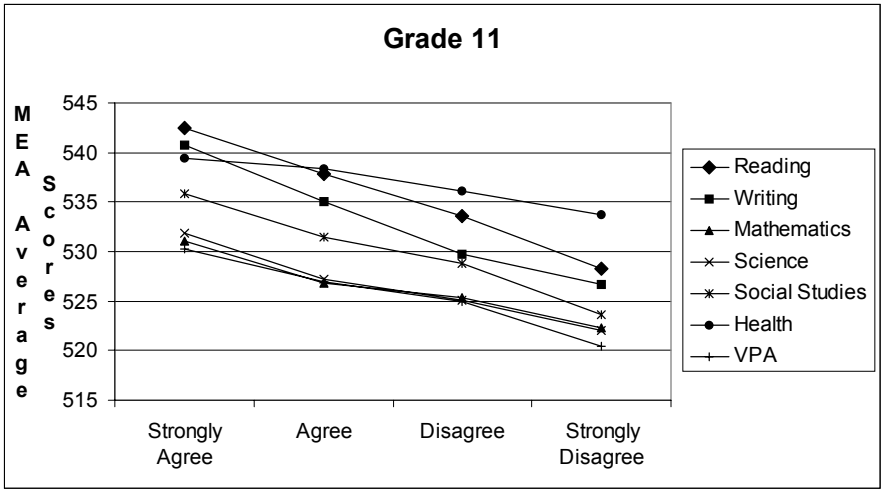
Questionnaire items related to examinees' attitudes toward different subjects tested in the MEA were administered to eighth- and eleventh-graders. For reading, writing, mathematics, science an technology, social studies, and visual and performing arts, students were asked how they feel about the statement, "My knowledge of [subject] will be useful to me in my future work." For health, students were asked how they feel about the statement, "My knowledge about health education will be helpful to me as an adult." Charts in Figure 15-3 indicate that students' attitudes toward the subjects tested in the MEA are related positively with MEA scores.

Figure 15-3

Attitude Towards Subject Matters and MEA Scores

Question: My knowledge of [subject] will be useful to me [in my future work/as an adult].





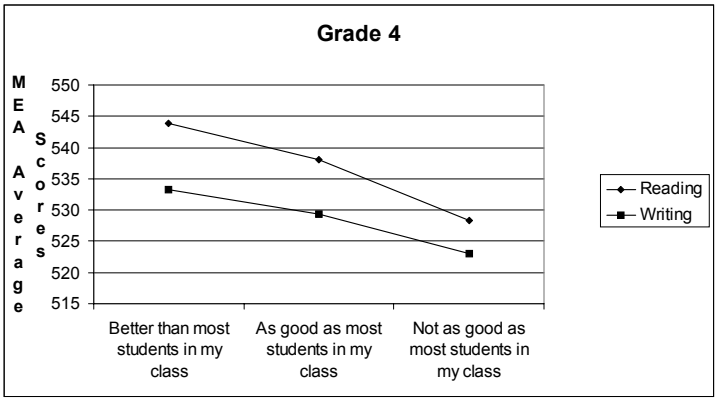
SELF IMAGE

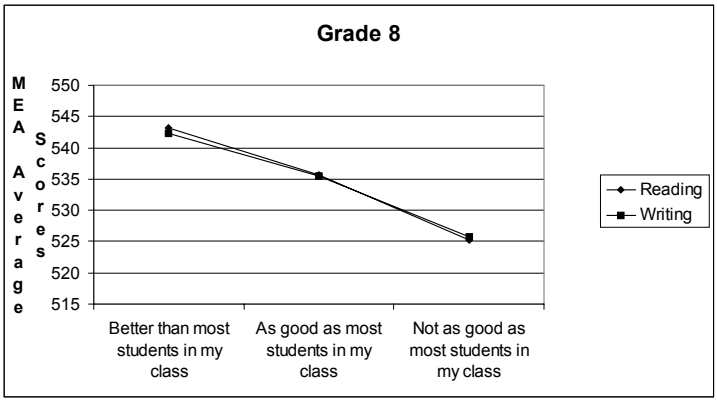
Students in all grades were asked, "How good are you at reading?" and, "How good are you at writing?" Figure 15-4 indicates that there is a positive relationship between students' self-image and their MEA scores in reading and writing.

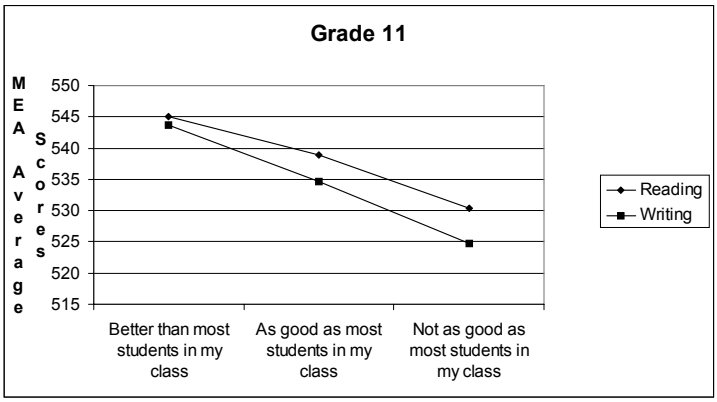
Figure 15-4

Self-Image and MEA Scores

Question: How good are you at reading/writing?





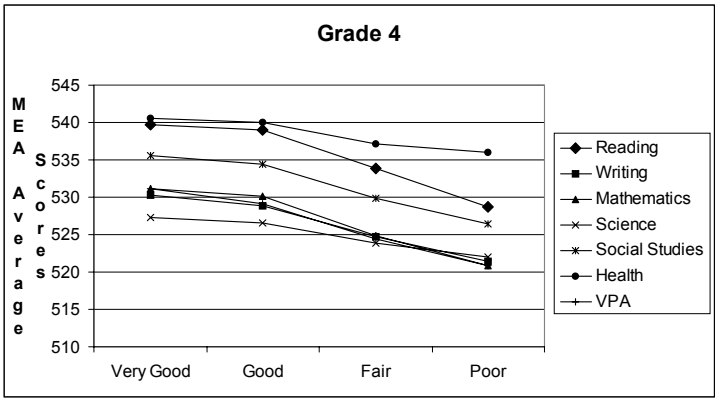


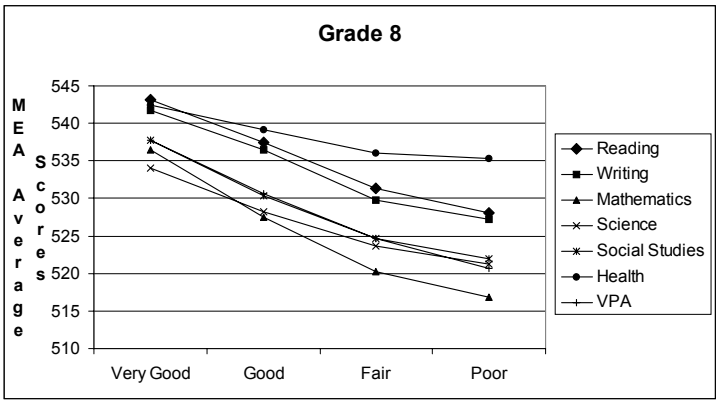
Students in grades 4 and 8 were asked, "Which of the following best describes how you rate yourself as a student?" Figure 15-5 indicates a positive relationship between self-image and MEA scores in all subject areas.

Figure 15-5

Self-Image and MEA Scores

Question: Which of the following best describes how you rate yourself as a student?





CHAPTER 16 — SCORE REPORTING

PRIMARY REPORTS

There were six primary reports for the 2001-02 MEA.

- Student Reports for Parent/Guardian
- Student Labels
- Common Item Level Class Report
- School Report
- District Report
- Student Writing CD

Except for the student writing CD, all of the primary reports listed are hardcopy reports. Each of these reports is described in the following subsections. Sample reports are provided in Appendix A.

STUDENT REPORT FOR PARENTS/GUARDIANS

The student report is a single-page report that is divided into three sections. The first section gives the student's overall performance for each content area. The student's scaled scores and performance levels are shown, both in a table and graphically. The graph shows the range of possible scaled scores, divided up into the four performance level ranges. For each content area, a diamond is printed in the appropriate location to show the student's scaled score, and a bar is printed around the diamond representing the standard error of measurement.

The second section of the student report compares the student's scores to the average scores for the school, district, and state. For each content area, a bar graph is printed that includes a bar for the student's scaled score and one for each of the three average scores included for comparison.

The third section of the report is a graph that shows the student's performance on the content area subcategories. The graph consists of a center line, which represents average performance for students who meet the standard, and three shaded bands. The innermost band marks off the area of

the graph that is within one standard deviation of the mean, the second band marks the area between one and two standard deviations from the mean, and the third is between two and three standard deviations from the mean. For each subcategory, the student's score is represented by a diamond printed in the appropriate place on the graph. (For a complete explanation of the content area subcategories, please see Chapter 12.) The report also includes definitions of the content area subcategories.

The reverse side of the student report provides a description of the performance levels and state summary results.

STUDENT LABELS

To aid schools in keeping track of student scores, schools were supplied with student score information on individual labels that they could affix to files, if desired.

COMMON ITEM CLASS REPORT

The common item class report provides a roster of all the students in each class and indicates their performance on the common items in the assessment. One report is provided for each content area. The student names are listed down the side of the report, and the item numbers are listed across the top. For each item, the following information is provided: the content standard measured by the item, the item type, the correct response (for multiple choice items) and the total possible points for the item. For each student, each multiple-choice item is marked either with a plus sign (+), indicating that the student chose the correct response, or a letter (A-D), indicating which incorrect response the student chose. For constructed-response items, the number of points the student attained is shown. At the end of the item responses, each student's total points earned, scaled score, and performance level are indicated. At the bottom of the report, the average percent correct on each item is shown for the class, school, district and state.

SCHOOL AND DISTRICT REPORTS

The school and district reports consist of three parts: the first part gives an overall summary of scores, the second provides a summary of student participation, and the third includes a report for each content area with more detailed scores.

The summary of scores includes a table that shows, for each content area, the average scaled score for the school, district, and state for each of the last three years, as well as a cumulative average across the three years. In addition, there is a bar graph for each content area that shows the percentage of students in each performance category at the school, district, and state levels. For the district version of this report, the school information is blank.

The summary of student participation gives the number and percentage of students who participated at the school, district, and state levels. These numbers are provided for the overall group of students as well as broken down by the following categories:

- ethnic group;
- whether the student has internet access at home;
- whether or not the student used accommodations and, for those who used accommodations, the reasons the accommodations were needed;
- students who were recommended for participation in the alternate assessment, reported overall as well as broken down by the reason for the use of the alternate assessment; and
- students who did not participate in part or all of the assessment

Again, for the district version of this report, the school information is blank.

For each content area, there is a two-page report showing results in more detail. The first page consists of two sections. The first section gives a definition of each of the performance levels along with a table showing the number and percentage of students at the school, district, and state who scored at each level for each of the past three years. The table also shows the cumulative

average over the three years. The second section provides results by the content area subcategories and the content standards. For each area, the table shows the total possible number of points and the average number and percent of points attained at the school, district and state levels. The school information is blank on the district-level reports.

The second page of the content area report shows results broken down by a number of different reporting categories (gender, ethnicity, internet access at home, Title 1 program, migrant, gifted/talented, disability, LEP status, and first grade of attendance in the district) as well as by responses to the questionnaire items. This information is provided for the school and the state on the school-level report and for the district and the state on the district-level report. For this table, results are only reported for groups with 5 or more students.

For each reporting category, the following information is given at the school or district level and at the state level:

- the percentage of students in that category
- the average scaled score for the group
- the percentage in the response category who meet or exceed the standard, partially meet the standard, and do not meet the standard.

For each questionnaire item response category, only the percentage of students in each category is reported at the school or district level. At the state level, the report shows the percentage of students in each category, the average scaled score, the percentage in the category who meet or exceed the standard, and the percentage who do not meet the standard.

DECISION RULES

To ensure that reported results for MEA 2001-2002 are accurate relative to collected data and other pertinent information, a document that delineates analysis and reporting rules was created.

These decision rules were observed in the analyses of MEA test data and in reporting the assessment results. Moreover, these rules are the main reference for quality assurance checks.

An excerpt of the decision rules document used for reporting results of the MEA December 2001 administration is in Appendix D. The same set of rules was used for reporting results of the MEA March 2002 administration, with adjustments made relative to the subjects tested.

The first set of rules pertains to general issues in reporting scores. Each issue is described and pertinent variables are identified. The actual rules applied are described by the way they impacts analyses and aggregations and their specific impact on each of the reports. The general rules are further grouped into issues pertaining to test items, school type, student exclusions, and number of students for aggregations.

The second set of rules pertains to reporting student participation. It describes which students were counted and reported for each subgroup in the student participation report.

QUALITY ASSURANCE

This section describes the different stages of the quality assurance program implemented for the MEA 2001-2002. The goals of the program are to

- ensure the accuracy of all data reported through independent verification of the calculated data.
- ensure all data reported are placed in the correct position on the report shell.
- ensure the report shell is grammatically and aesthetically correct.

Checklists that were used in the quality assurance process for MEA are included in Appendix E.

STAGE 1

The MEA Quality Assurance Program commences once the following occurs:

1. The data analyst accepts the raw test data results from Data Processing.

2. The report shells have been updated, quality reviewed, and approved by the DOE.
3. The Decision Rules, including calculation methods, have been documented and approved by the DOE.

STAGE 2

Reference information is collected prior to and during the review process, including

1. District, School and Class names, census, and codes
2. List of students who are reporting exceptions
3. List of home-schooled students
4. Proficiency level scaled score ranges
5. Answer keys, item types, and item categories for sub score reporting
6. Raw score to scaled score conversion tables
7. DOE approved state results

STAGE 3

Review the decision rules for any unique reporting situations and, using the district, school, and class list, select a sample of districts and schools for the QA review, being sure to include districts/schools with unique reporting requirements.

STAGE 4

Score the test for each student. The following steps are completed for each content area.

1. Copy the file from Data Processing with the test results for each student to an excel spreadsheet.
2. Using the item information, score the common items for each child; that is, replace all correct answers with a "1".
3. Compute the raw score for each student by adding up the "1's" for each student.
4. Using the conversion table and the raw score, determine the scaled score and performance level for each student.

5. Using the Decision Rules, remove to a separate spreadsheet all students exempted from reporting. Compare to the lists of exempted students and investigate any differences.

STAGE 5

Compute and verify the state average percent correct for each common item.

STAGE 6

Compute and verify the state Average Performance Score.

STAGE 7

Compute and verify state counts on the Summary of Student Participation page.

STAGE 8

Compute and verify the state performance level percentages.

STAGE 9

Compute state averages and percentages for reporting categories and questionnaire items.

STAGE 10

Using the list of sample districts previously selected, copy the students for each sample district to a separate worksheet. Compute the same averages and percents for the school and district level as in steps E – I above.

STAGE 11

Print all the common item, school, and district reports, labels, and a sample of student reports for the sample districts. Using the above computed data in conjunction with the attached check off sheet for each report or file, review the report output. If problems are found two things are done:

1. Advise the Report Programmers or the Data Analyst if there is a problem.
2. Document the problem and follow up and verify the correction was made.

STAGE 12

When all corrections have been made and QA staff is satisfied that the reports are correct, move a copy of the report files to the appropriate folder in FINAL REPORTS and advise that the files may be sent for printing.

REFERENCES

Allen, Mary J. & Yen, Wendy M. (1979). *Introduction to Measurement Theory.* Belmont, CA: Wadsworth, Inc.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd Edition). Fort Worth: Holt, Rinehart and Winston.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description. In P. W. Holland & H. Wainer (Eds.) *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23, 355-368.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.) *Educational Measurement* (3rd Ed. Pp. 105-146).

Holland, P.W., & Wainer, H. (1993). *Differential item functioning.* Hillsdale, NJ: Erlbaum.

Joint Committee on Testing Practices (1988). *Code of Fair Testing Practices in Education*. Washington, D.C.: National Council on Measurement in Education.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179-197.

Lord, F.M. & Novick, M.R. (1968). Statistical Theories of Mental Test Score*s.* Reading, MA: Addison-Wesley.

Medsker, G. J. (1998*). Determining the relationship between student transience and KIRIS school results: are schools with transient students unfairly impacted?* (HumRRO Report, FR-WATSD-98-12). Alexandria, VA: Human Resources Research Organization.

Petersen, N.S., Kolen, M.J., & Hoover, H.D. (1989) Scaling, norming, and equating. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221-262).

APPENDIX A
SAMPLE REPORTS

Maine Educational Assessment

Common Item Class Report

ELA READING

Grade 11

Code:
District:
School:
Class:
Date: December 2001
Group Size:
Page: of

Name	Item Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	Points Earned (52 Max. Points)	Scaled Score	Performance Level
	Content Standard and Performance Indicator	A3	D2	D5	C1	D5	D5	D5	A3	D5	D5	D5	C1	B2	C1	B2	B7	B7	D5	A5	D5	A3	D5	D5	B7	B6	A6	B7	A3	C1	B7	B8	B5	B7			
	Item Type	MC	MC	MC	MC	MC	MC	MC	MC	SA	CR	CR	MC	MC	MC	MC	SA	CR	MC	MC	MC	MC	SA	CR	MC	MC	MC	MC	MC	MC	MC	MC	SA	ER			
	Correct MC Response	B	D	C	A	B	D	D	A				C	B	D	A			C	D	C	A			A	A	B	B	C	B	D	A					
	Total Possible Points	1	1	1	1	1	1	1	1	2	4	4	1	1	1	1	2	4	1	1	1	1	2	4	1	1	1	1	1	1	1	1	2	4			

Item Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Percent Correct/Avg. Score: Class																																	
Percent Correct/Avg. Score: School																																	
Percent Correct/Avg. Score: District																																	
Percent Correct/Avg. Score: State																																	

Common Item Class Report

ELA WRITING

Grade 11

Code:
District:
School:
Class:
Date: December 2001
Group Size: Page: of

Name	Writing Prompt			Extended-Response/Writing Sample			Total Writing			Scaled Score	Performance Level
	Stylistic and Rhetorical Aspects (12 possible points)	Standard English Conventions (8 possible points)	Total (20 possible points)	Stylistic and Rhetorical Aspects (6 possible points)	Standard English Conventions (4 possible points)	Total (10 possible points)	Stylistic and Rhetorical Aspects (18 possible points)	Standard English Conventions (12 possible points)	Total (30 possible points)		
Percent Correct/Avg. Score: Class											
Percent Correct/Avg. Score: School											
Percent Correct/Avg. Score: District											
Percent Correct/Avg. Score: State											

Maine Educational Assessment

Name:
School:

District:
------------------ **Performance Levels** ------- **Scaled Scores**

Grade: **Writing:**
Date: 12/01 **Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* required by law to be fully implemented by 2002-2003.

Maine Educational Assessment

Name:
School:

District:
------------------ **Performance Levels** ------- **Scaled Scores**

Grade: **Writing:**
Date: 12/01 **Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* required by law to be fully implemented by 2002-2003.

Maine Educational Assessment

Name:
School:

District:
------------------ **Performance Levels** ------- **Scaled Scores**

Grade: **Writing:**
Date: 12/01 **Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* required by law to be fully implemented by 2002-2003.

Maine Educational Assessment

Name:
School:

District:
------------------ **Performance Levels** ------- **Scaled Scores**

Grade: **Writing:**
Date: 12/01 **Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* required by law to be fully implemented by 2002-2003.

Maine Educational Assessment

Name:
School:

District:
------------------ **Performance Levels** ------- **Scaled Scores**

Grade: **Writing:**
Date: 12/01 **Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* required by law to be fully implemented by 2002-2003.

Maine Educational Assessment

Name:
School:

District:
------------------ **Performance Levels** ------- **Scaled Scores**

Grade: **Writing:**
Date: 12/01 **Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* required by law to be fully implemented by 2002-2003.

Maine Educational Assessment

Name:
School:

District:
------------------ **Performance Levels** ------- **Scaled Scores**

Grade: **Writing:**
Date: 12/01 **Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* required by law to be fully implemented by 2002-2003.

Maine Educational Assessment

Name:
School:

District:
------------------ **Performance Levels** ------- **Scaled Scores**

Grade: **Writing:**
Date: 12/01 **Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* required by law to be fully implemented by 2002-2003.

Maine Educational Assessment

Name:
School:

District:
------------------ **Performance Levels** ------- **Scaled Scores**

Grade: **Writing:**
Date: 12/01 **Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* required by law to be fully implemented by 2002-2003.

Maine Educational Assessment

Name:
School:

District:
------------------ **Performance Levels** ------- **Scaled Scores**

Grade: **Writing:**
Date: 12/01 **Reading:**

The MEA was revised in 1998/99 to assess Maine's *Learning Results,* required by law to be fully implemented by 2002-2003.

Important Information for Parents/Guardians Grade 11 Assessment December 2001 Administration

STATE OF MAINE
DEPARTMENT OF EDUCATION
23 State House Station
Augusta, ME 04333
June 2002

J. Duke Albanese
COMMISSIONER

Dear Parents/Guardians:

The Legislature approved Maine's *Learning Results* in May 1997, giving all schools standards to measure student learning. Our goal is for all students in Maine to demonstrate that they meet all standards defined in the *Learning Results.* The Maine Educational Assessment (MEA), the state test that has been administered for the past 17 years in grades 4, 8, and 11, has been rewritten so it tests the challenging subject matter specified in the *Learning Results.* Your student was in the fourth group to take this revised test. In the past, the MEA reported individual student scores only in reading and writing (called English Language Arts in the *Learning Results*) and in mathematics. The revised MEA also includes individual results in science and technology and in social studies.

During December 2001, students in grades 4, 8, and 11 were tested in reading and writing as part of the MEA. The test included a composition, multiple-choice questions, short-answer questions, and essay questions (known as "constructed response"). The report on the reverse side of this letter provides you with important information about your student's performance on this part of the MEA, along with a summary of school, district, and state results. Please keep in mind that your student's score measures learning over the past 3–4 years, not just the work of the past year.

In March 2002 your student participated in the fourth administration of the revised MEA in the subjects of mathematics, science and technology, and social studies. We anticipate that the results of the March testing will be available in September.

Staff at your school can provide further information about school and district results as well as about your student's overall performance. The MEA is just one part of the comprehensive assessment system your student's school uses to measure student learning and school success. MEA results are used at the school, district, and state levels to improve teaching and learning.

Sincerely,

J. Duke Albanese
Commissioner

Information on Maine's *Learning Results*

- The *Learning Results* were developed in eight content areas by thousands of Maine citizens.
- The MEA was rewritten by hundreds of Maine teachers to align with the *Learning Results.*
- Setting MEA performance standards based on the quality of student work was completed by hundreds of Maine teachers and citizens.
- For a copy of Maine's *Learning Results,* either call 624-6629 or find them on-line at http://www.state.me.us/education/lres/homepage.htm.

Performance Levels and Score Ranges

On this assessment, results are reported as four performance levels using scaled scores that range from 501 to 580. **The text below describes the quality of student work for each performance level.**

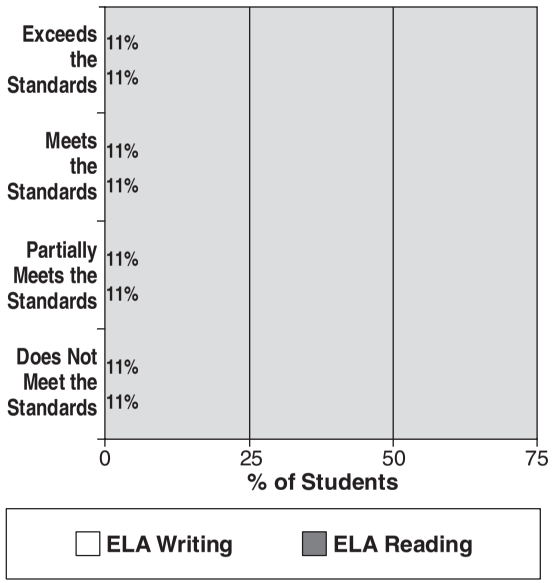
❒ **Exceeds the Standards (561 to 580)**
The student's work demonstrates exemplary accomplishment of content knowledge, analysis, problem-solving, and communication skills.

❒ **Meets the Standards (541 to 560)**
The student's work demonstrates consistent accomplishment of content knowledge, analysis, problem-solving, and communication skills.

❒ **Partially Meets the Standards (521 to 540)**
The student's work demonstrates inconsistent accomplishment of content knowledge, analysis, problem-solving, and communication skills.

❒ **Does Not Meet the Standards (501 to 520)**
The student's work demonstrates limited command of content knowledge, analysis, problem-solving, and communication skills.

Maine State MEA Summary Results for December 2001 Administration



Student	Grade	School	District
	11		

Content Area	Performance Level	Score	This Student's Performance Levels and Scores
			Does Not Meet the Standards / Partially Meets the Standards / Meets the Standards / Exceeds the Standards
ELA* Writing			
ELA* Reading			

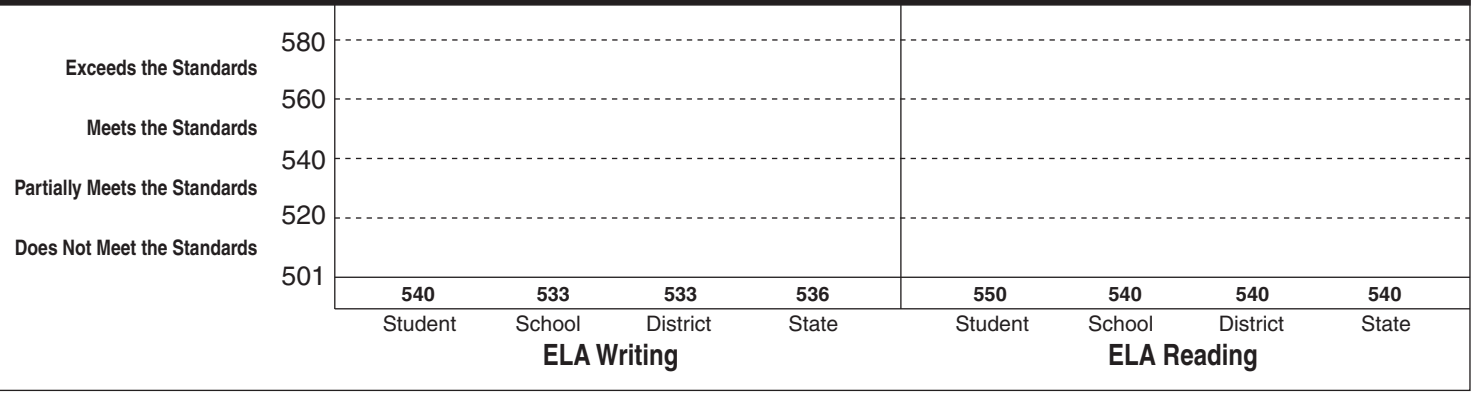
501 520 540 560 580

***ELA** is an abbreviation for English Language Arts
Testing Incomplete (TI): Student failed to attempt one or more sessions.

See reverse side for description of performance levels and state summary results.

The diamond (◆) represents the student's score. The bar (▬) surrounding the score represents the probable range of scores for the student if he or she were to be tested many times. This statistic is called the standard error of measurement.

How this Student's Performance Compared to Average Scores from School, District, and State



This Student's Performance in Content Area Subcategories

Content Areas	Content Area Subcategories	Student's Score Compared with Meeting the State Standards (Weaker / Meets the Standards / Stronger)
ELA Writing	Standard English Conventions (Standard F)	
	Stylistic and Rhetorical Aspects of Writing (Standard G)	
ELA Reading	Reading Process, Language, and Comprehension (Standards A, B, C, D)	

Definitions of Content Area Subcategories

Standard English Conventions: Refers to a student's ability to write correctly. Scoring focused on sentence structure, grammar and usage, and mechanics.

Stylistic and Rhetorical Aspects of Writing: Refers to a student's ability to use writing to explore ideas, to present lines of thought, to represent and reflect on human experience, and to communicate feelings, knowledge, and opinions. Scoring focused on topic development, organization, use of supportive details, and varied language and style.

Reading Process, Language, and Comprehension: Refers to a student's level of comprehension of literary reading selections (e.g., fiction, short stories, poetry) and informational reading selections (e.g., newspaper articles, informational essays, textbook passages), as well as a student's use of reading strategies, language, and analysis.

DEPARTMENT OF EDUCATION

2001-2002 School Year Reports

Dear School Board Members and School Personnel:

The Maine Educational Assessment (MEA) is in its fourth year of measuring student performance on Maine's *Learning Results*, which challenge schools and students to pursue academic standards that are among the highest in the nation. This report of student performance in reading, writing, and health education on tests administered in December 2001 is the first of two summary reports you will be receiving for the school year 2001 through 2002. The second report will include results from the assessment of mathematics, science and technology, social studies, and visual and performing arts administered in March of 2002. This MEA results report should still be considered baseline information, as the *Learning Results* are not scheduled for full implementation until the 2003 school year.

The MEA, revised to align with Maine's *Learning Results*, is composed of selected-response (multiple-choice) questions, as well as short-answer questions and complex questions, including a writing prompt, that require students to construct answers that demonstrate their knowledge and skills. Your review of the MEA questions that we have released will help you understand the revised assessment and the challenge that it presents for students and schools. The scores are reported using a numerical scale (501–580) and the performance levels "Does Not Meet the Standards," "Partially Meets the Standards," "Meets the Standards," and "Exceeds the Standards." The scale and the performance levels, established in the fall of 1999, will remain fixed for a period of at least five years to measure progress of students across the state in achieving the standards.

It is important to know that more than 500 teachers and other educators from across Maine helped to develop the revised MEA and assisted in the scoring and standard-setting process. Maine teachers continue to advise annually the updating and development of MEA tests. This participation has not only strengthened the redesigned MEA but has also engaged teachers from around the state in conversations about quality standards for student work.

I look forward to continuing the strong local and state partnership that has led to our current success as we work toward achieving even higher standards for all Maine students.

Sincerely,

J. Duke Albanese
Commissioner

MAINE EDUCATIONAL ASSESSMENT

Educational Assessment School Report

ID:

School:

District:

Grade: 11

Test Date: DECEMBER 2001

Contents of the Report

The report is divided into five main sections including a section describing the students tested and a separate section for the results in each content area.

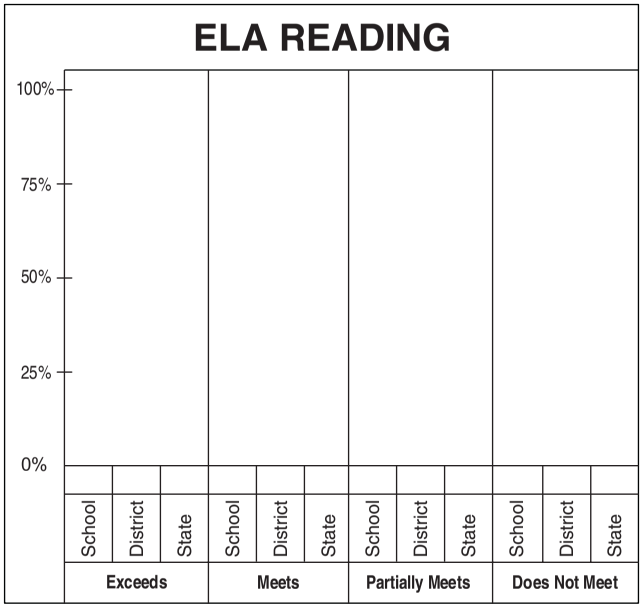
MAINE EDUCATIONAL ASSESSMENT

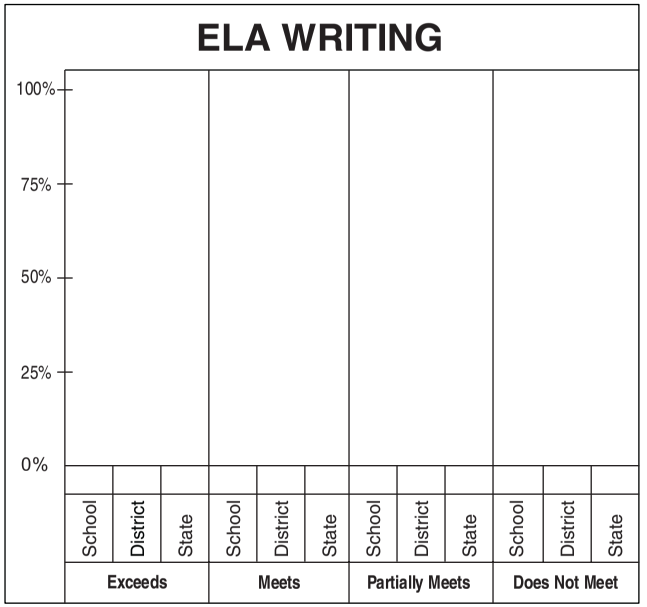
SUMMARY OF SCORES

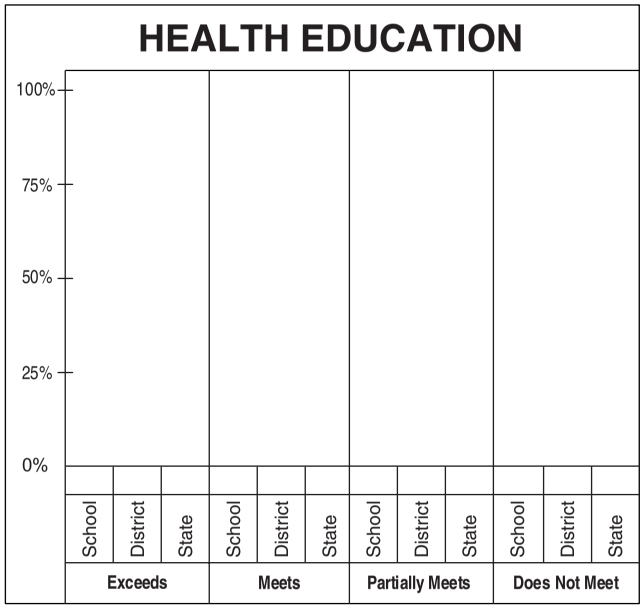
School:
District:
Grade: 11
Date: DECEMBER 2001

Executive Summary of School, District, and State Scores

Year	Average Performance Score		
	School	District	State
ELA READING			
1999–2000			
2000–2001			
2001–2002			
Cum. Avg.			
ELA WRITING			
1999–2000			
2000–2001			
2001–2002			
Cum. Avg.			
HEALTH EDUCATION			
1999–2000			
2000–2001			
2001–2002			
Cum. Avg.			







SUMMARY OF STUDENT PARTICIPATION

School:
District:
Grade: 11
Date: DECEMBER 2001

Participation Category	Number			Percentage		
	State	District	School	State	District	School
Students enrolled on the first day of testing						
Ethnicity						
White (non-Hispanic)						
Black (non-Hispanic)						
Hispanic						
Asian/Pacific Islander						
American Indian/Alaskan Native						
Multi-ethnic						
Other						
Not reported						
Internet access at home						
Yes						
No						
Students who took all or part of the assessment without accommodations						
Students who took all or part of the assessment with accommodations						
Identified disability (PET/IEP)						
LEP						
504 plan						
Other						
Reason not reported						
Students recommended for participation in alternate assessment (PAAP)						
Identified disability (PET/IEP)						
LEP						
504 plan						
Reason not reported						
Students who did not participate in all or part of the assessment due to absence						
Students who did not participate in all or part of the assessment due to other reasons						

MAINE EDUCATIONAL ASSESSMENT

ELA READING RESULTS

School:
District:
Grade: 11
Date: DECEMBER 2001

PERFORMANCE LEVELS	STUDENTS AT EACH PERFORMANCE LEVEL					
		School		District		State
Exceeds the Standards—The quality of a student's work at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The work demonstrates exemplary accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate (scaled scores: 561–580).	1999–2000 2000–2001 **2001–2002** Cumulative Average					
Meets the Standards—The quality of a student's work at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The work demonstrates a consistent accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate (scaled scores: 541–560).	1999–2000 2000–2001 **2001–2002** Cumulative Average					
Partially Meets the Standards—The quality of a student's work at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The work demonstrates inconsistent accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate (scaled scores: 521–540).	1999–2000 2000–2001 **2001–2002** Cumulative Average					
Does Not Meet the Standards—The quality of a student's work at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The student demonstrates limited accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate (scaled scores: 501–520).	1999–2000 2000–2001 **2001–2002** Cumulative Average					

Learning Results Content Standards	Number of Points Possible	Average Points Attained (Number and Percent)					
		School		District		State	
		N	%	N	%	N	%
Reading Process and Language (Standards A and C)							
Reading Comprehension (Standards B and D)							
Literature and Culture (Standard B)							
Informational Texts (Standard D)							

MAINE EDUCATIONAL ASSESSMENT

ELA READING RESULTS
(CONTINUED)

School:
District:
Grade: 11
Date: DECEMBER 2001

Reporting Categories	School					State				
	% Students in Each Category	Scaled Score	% Exceeds or Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards	% Students in Each Category	Scaled Score	% Exceeds or Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards
Gender										
female										
male										
Ethnicity										
White (non-Hispanic)										
Black (non-Hispanic)										
Hispanic										
Asian/Pacific Islander										
American Indian/Alaskan native										
multi-ethnic										
other										
not reported										
Internet access at home										
yes										
no										
Migrant										
students eligible, not served										
students eligible, served, not tutored										
students eligible, served, tutored										
Gifted/talented program										
yes										
no										
Identified disability										
yes										
no										
Language minority/LEP student										
bilingual never identified LEP										
former LEP reclassified non-LEP										
current LEP										
First grade in district										
before grade 9										
grade 9										
grade 10										
grade 11										
College prep										
yes										
no										
Parent education										
did not finish high school										
graduated from high school										
some education after high school										
college and/or advanced degree										

Questionnaire Items	Sch.	State			
	% Students in Each Category	% Students in Each Category	Scaled Score	% Exceeds or Meets the Standards	% Does Not Meet the Standards
How often are you asked to do research using information from one or more content areas?					
once a week					
at least once a month					
at least once a term					
never					
How many books have you read at home in the past two months?					
none					
one					
two to four					
five or more					
How often do you search for and read information on a computer?					
several times a week					
once a week					
at least once a month					
never					
How do you feel about the following statement? "My knowledge of reading will be useful to me as an adult."					
strongly agree					
agree					
disagree					
strongly disagree					
How good are you at reading?					
I am better than most students in my class.					
I am as good as most students in my class.					
I am not as good as most students in my class.					
High school career pathway					
college prep					
tech prep					
occupational prep					
apprenticeship programs					
Hours worked at part-time job during school week					
do not work part-time during school week					
8 hours or fewer					
9-21 hours					
more than 21 hours					

MAINE EDUCATIONAL ASSESSMENT

ELA WRITING RESULTS

School:
District:
Grade: 11
Date: DECEMBER 2001

PERFORMANCE LEVELS	STUDENTS AT EACH PERFORMANCE LEVEL					
		School		District		State
Exceeds the Standards—The quality of a student's written compositions at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates exemplary accomplishment in both the development of the topic/idea and the use of Standard English conventions in first-draft writing (scaled scores: 561–580).	1999–2000 2000–2001 **2001–2002** Cumulative Average					
Meets the Standards—The quality of a student's written compositions at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates proficiency in both the development of the topic/idea and the use of Standard English conventions in first-draft writing (scaled scores: 541–560).	1999–2000 2000–2001 **2001–2002** Cumulative Average					
Partially Meets the Standards—The quality of a student's written compositions at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates writing skills that may show moderate development of the topic/idea and/or some errors in Standard English conventions that may interfere with communication of ideas (scaled scores: 521–540).	1999–2000 2000–2001 **2001–2002** Cumulative Average					
Does Not Meet the Standards—The quality of a student's written compositions at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates writing skills that show limited development of the topic/idea and/or many errors in Standard English conventions that interfere with communication of ideas (scaled scores: 501–520).	1999–2000 2000–2001 **2001–2002** Cumulative Average					

Learning Results Content Standards	Number of Points Possible	Average Points Attained (Number and Percent)					
		School		District		State	
		N	%	N	%	N	%
Writing (Standards F and G)							
Standard English Conventions (Standard F)							
Stylistic and Rhetorical Aspects of Writing (Standard G)							

MAINE EDUCATIONAL ASSESSMENT

ELA WRITING RESULTS
(CONTINUED)

School:
District:
Grade: 11
Date: DECEMBER 2001

Reporting Categories	School					State				
	% Students in Each Category	Scaled Score	% Exceeds or Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards	% Students in Each Category	Scaled Score	% Exceeds or Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards
Gender										
female										
male										
Ethnicity										
White (non-Hispanic)										
Black (non-Hispanic)										
Hispanic										
Asian/Pacific Islander										
American Indian/Alaskan native										
multi-ethnic										
other										
not reported										
Internet access at home										
yes										
no										
Migrant										
students eligible, not served										
students eligible, served, not tutored										
students eligible, served, tutored										
Gifted/talented program										
yes										
no										
Identified disability										
yes										
no										
Language minority/LEP student										
bilingual never identified LEP										
former LEP reclassified non-LEP										
current LEP										
First grade in district										
before grade 9										
grade 9										
grade 10										
grade 11										
College prep										
yes										
no										
Parent education										
did not finish high school										
graduated from high school										
some education after high school										
college and/or advanced degree										

Questionnaire Items	Sch.	State			
	% Students in Each Category	% Students in Each Category	Scaled Score	% Exceeds or Meets the Standards	% Does Not Meet the Standards
How much in-school time do you spend writing each week?					
less than 45 minutes					
about an hour					
1 1/2 to 2 hours					
2 1/2 hours or more					
How do you use a computer for writing?					
not at all					
drafts only					
drafts and final copy					
final copy only					
Do you or your teacher keep a collection of your writing?					
A collection of my writing is not kept.					
A collection of my writing is kept, but I don't use it.					
A collection of my writing is kept and I use it to grow as a writer.					
How do you most often receive grammar instruction?					
individually, during writing conferences					
by written comments on my papers					
in mini-lessons during English class					
in a separate class based on a grammar textbook					
How good are you at writing?					
I am better than the average student in my classes.					
I am as good as the average student in my classes.					
I am not as good as the average student in my classes.					
How do you feel about the following statement? "My ability to write will be useful to me as an adult."					
strongly agree					
agree					
disagree					
strongly disagree					
High school career pathway					
college prep					
tech prep					
occupational prep					
apprenticeship programs					
Hours worked at part-time job during school week					
do not work part-time during school week					
8 hours or fewer					
9-21 hours					
more than 21 hours					

MAINE EDUCATIONAL ASSESSMENT

HEALTH EDUCATION RESULTS

School:
District:
Grade: 11
Date: DECEMBER 2001

PERFORMANCE LEVELS	STUDENTS AT EACH PERFORMANCE LEVEL	School		District		State
Exceeds the Standards—The quality of a student's body of work at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates exemplary knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis, and risk reduction (scaled scores: 561–580).	1999–2000 2000–2001 **2001–2002** Cumulative Average					
Meets the Standards—The quality of a student's body of work at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates consistent knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis, and risk reduction (scaled scores: 541–560).	1999–2000 2000–2001 **2001–2002** Cumulative Average					
Partially Meets the Standards—The quality of a student's body of work at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates partial and/or inconsistent knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis, and risk reduction (scaled scores: 521–540).	1999–2000 2000–2001 **2001–2002** Cumulative Average					
Does Not Meet the Standards—The quality of a student's body of work at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates a limited knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis, and risk reduction (scaled scores: 501–520).	1999–2000 2000–2001 **2001–2002** Cumulative Average					

Learning Results Content Standards	Number of Points Possible	Average Points Attained (Number and Percent)					
		School		District		State	
		N	%	N	%	N	%
Health Concepts (Standard A)							
Health Information, Services, and Products (Standard B)							
Health Promotion and Risk Reduction (Standard C)							
Influences on Health (Standard D)							
Communication Skills (Standard E)							
Decision Making and Goal Setting (Standard F)							
Community, Consumer, and Environmental Health							
Personal and Nutritional Health							
Family Life Education and Growth and Development							
Safety and Injury Prevention							
Tobacco, Alcohol, and Other Drug Use Prevention							
Prevention and Control of Disease and Disorders							

HEALTH EDUCATION RESULTS
(CONTINUED)

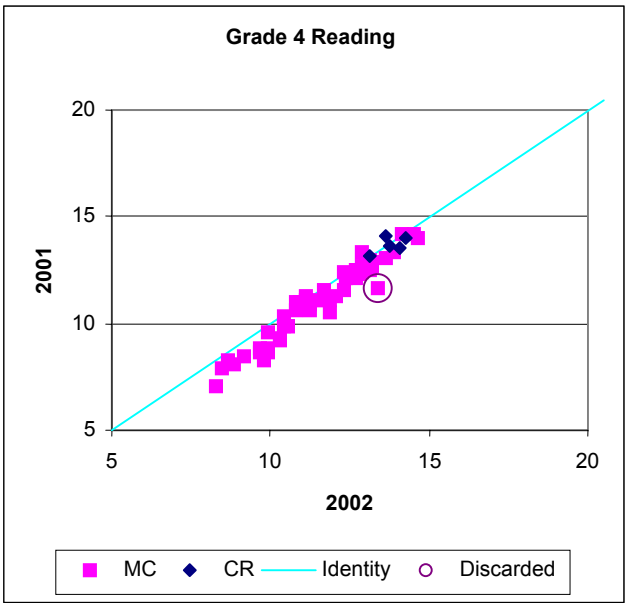
School:
District:
Grade: 11
Date: DECEMBER 2001

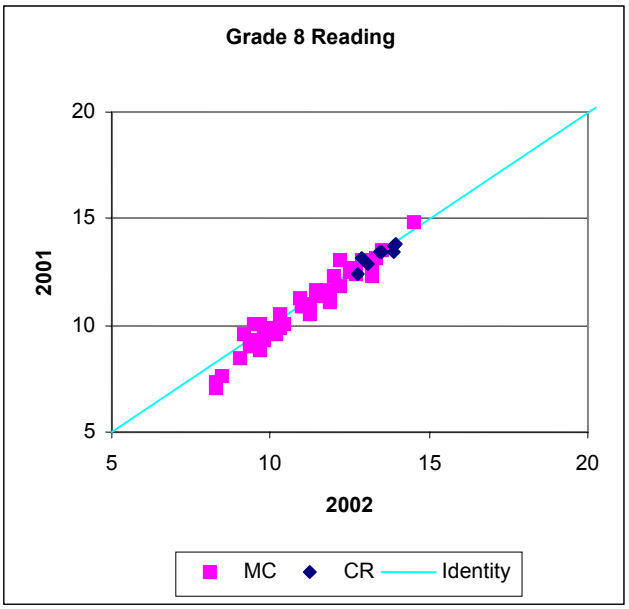
Reporting Categories	School					State				
	% Students in Each Category	Scaled Score	% Exceeds or Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards	% Students in Each Category	Scaled Score	% Exceeds or Meets the Standards	% Partially Meets the Standards	% Does Not Meet the Standards
Gender										
female										
male										
Ethnicity										
White (non-Hispanic)										
Black (non-Hispanic)										
Hispanic										
Asian/Pacific Islander										
American Indian/Alaskan native										
multi-ethnic										
other										
not reported										
Internet access at home										
yes										
no										
Migrant										
students eligible, not served										
students eligible, served, not tutored										
students eligible, served, tutored										
Gifted/talented program										
yes										
no										
Identified disability										
yes										
no										
Language minority/LEP student										
bilingual never identified LEP										
former LEP reclassified non-LEP										
current LEP										
First grade in district										
before grade 9										
grade 9										
grade 10										
grade 11										
College prep										
yes										
no										
Parent education										
did not finish high school										
graduated from high school										
some education after high school										
college and/or advanced degree										

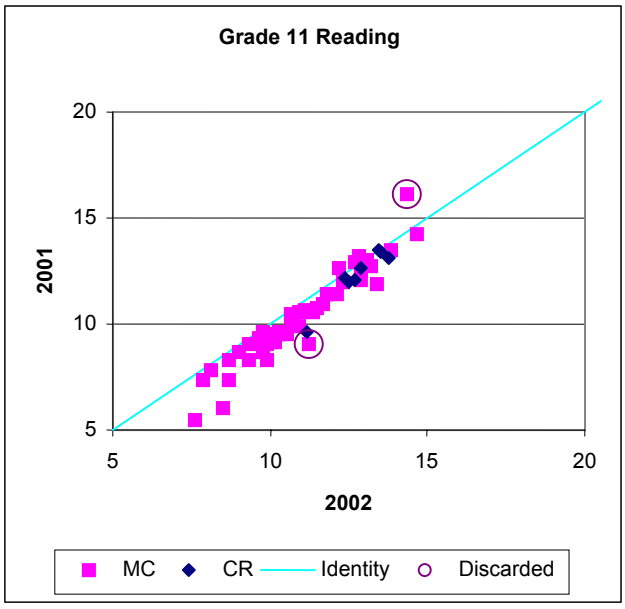
Questionnaire Items	Sch.	State			
	% Students in Each Category	% Students in Each Category	Scaled Score	% Exceeds or Meets the Standards	% Does Not Meet the Standards
How much did you learn about predicting the immediate and long-term impact of health decisions in your high school health education class?					
a lot					
some					
nothing					
How much did you learn about the relationship between health practices and individual well-being in your high school health education class?					
a lot					
some					
nothing					
How do you feel about the following statement? "My knowledge of health education will be useful to me as an adult."					
strongly agree					
agree					
disagree					
strongly disagree					
Think about what you learned in high school health education class. Which area have you found most useful?					
growth and development, such as physical changes; and personal hygiene including physical activity					
mental health, such as stress management					
nutrition, such as eating healthy snacks					
substance abuse prevention, such as tobacco, alcohol, and other drugs					
High school career pathway					
college prep					
tech prep					
occupational prep					
apprenticeship programs					
Hours worked at part-time job during school week					
do not work part-time during school week					
8 hours or fewer					
9-21 hours					
more than 21 hours					

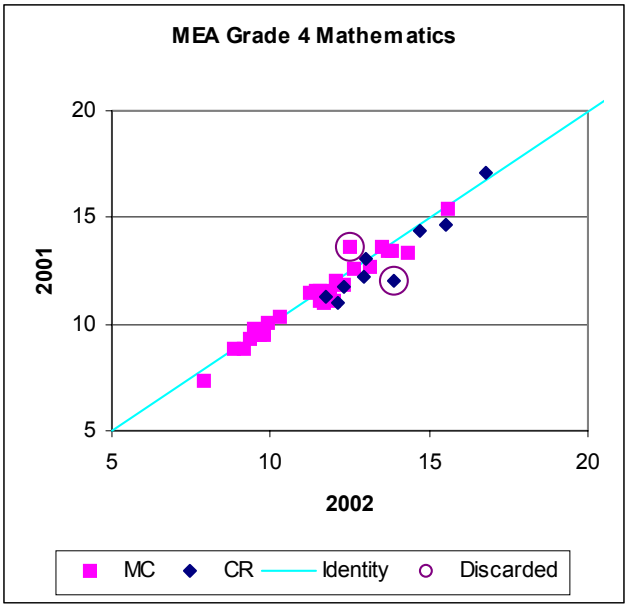
APPENDIX B

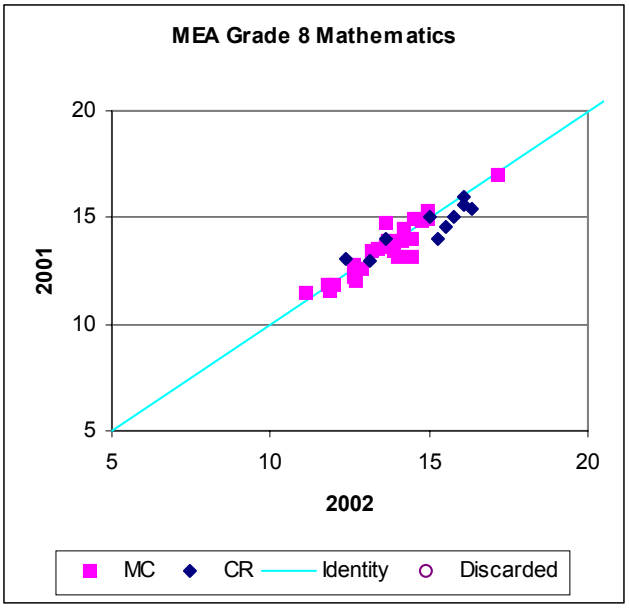
DELTA PLOTS

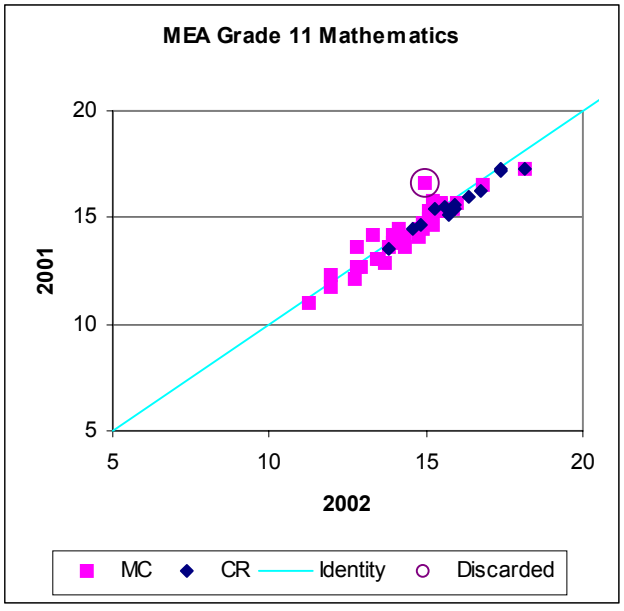

Grade 4 Reading
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

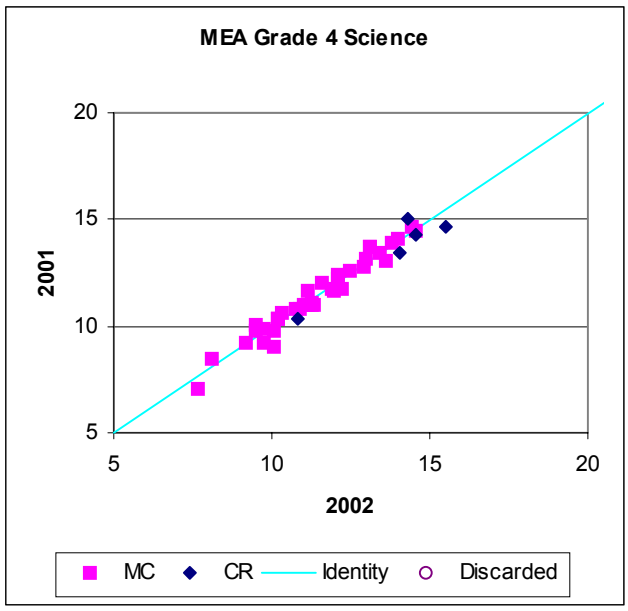

Grade 8 Reading
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity

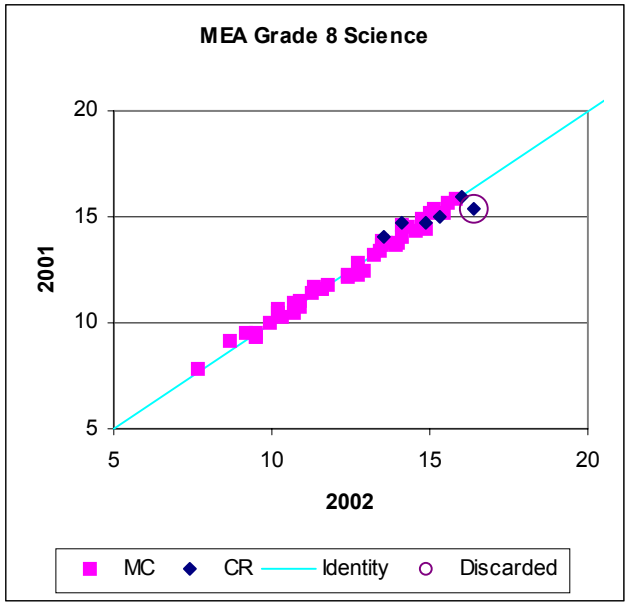

Grade 11 Reading
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

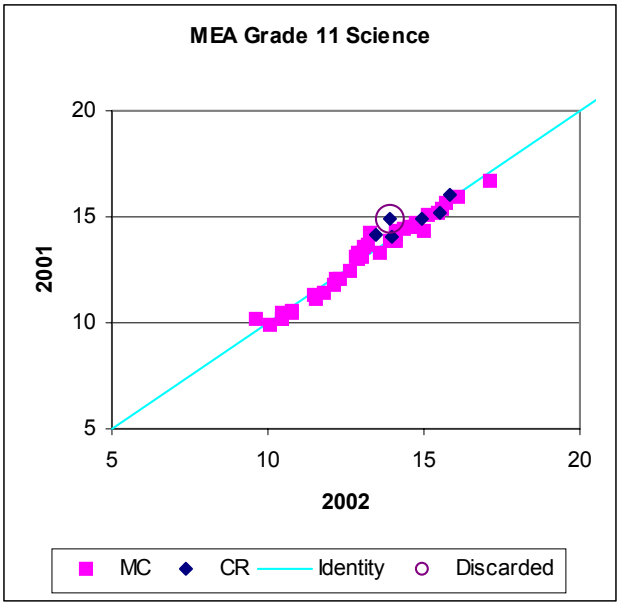

MEA Grade 4 Mathematics
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

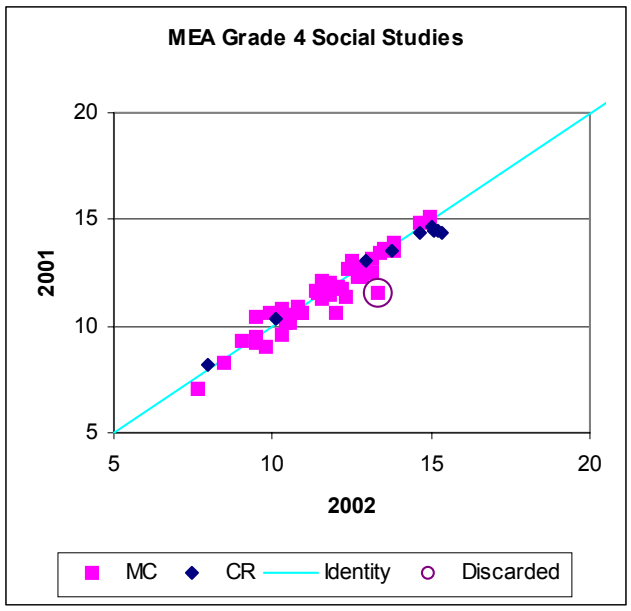

MEA Grade 8 Mathematics
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

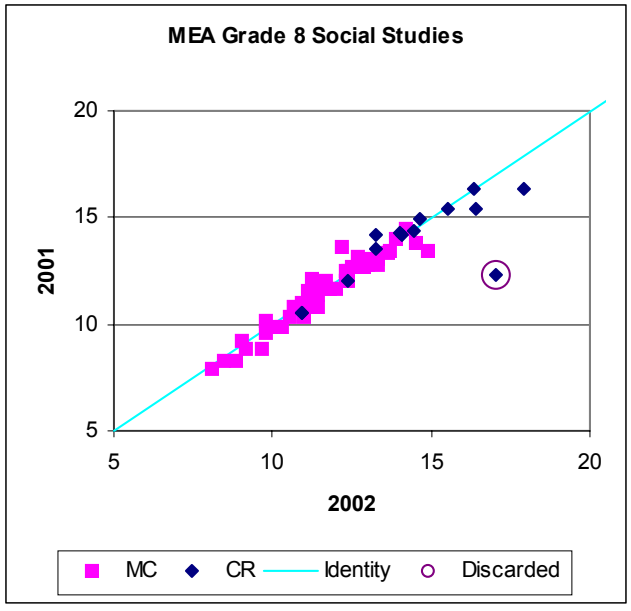

MEA Grade 11 Mathematics
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

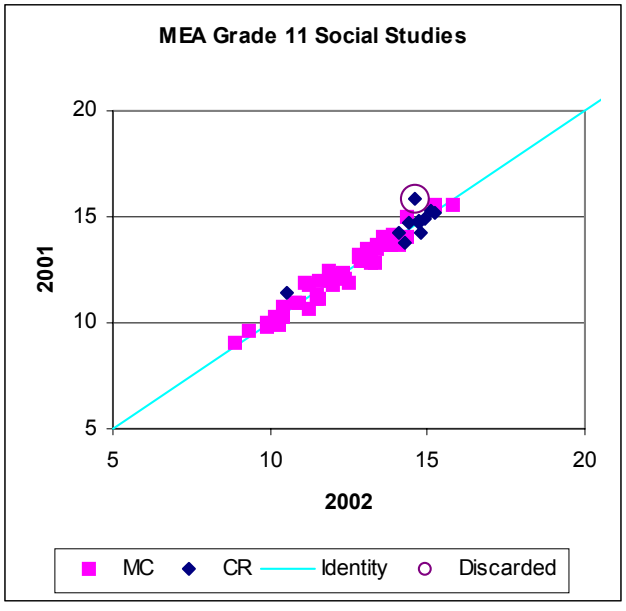

MEA Grade 4 Science
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

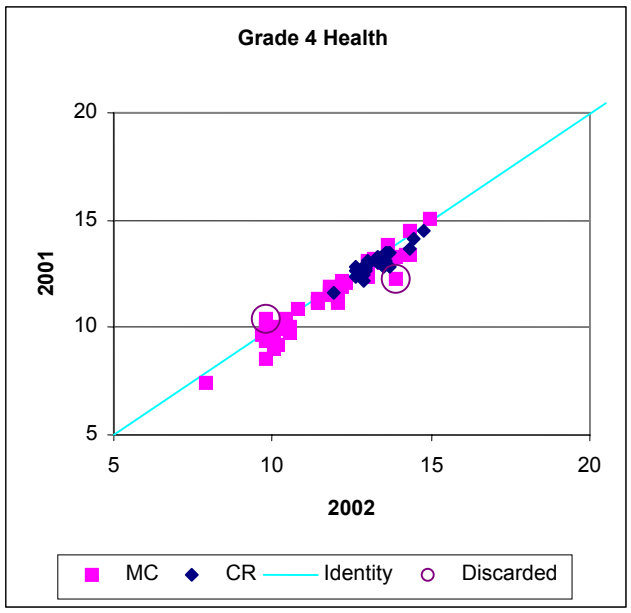

MEA Grade 8 Science
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

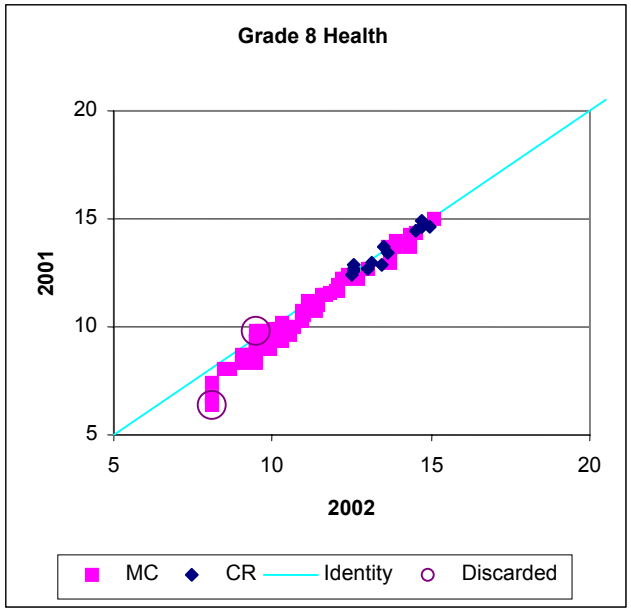

MEA Grade 11 Science
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

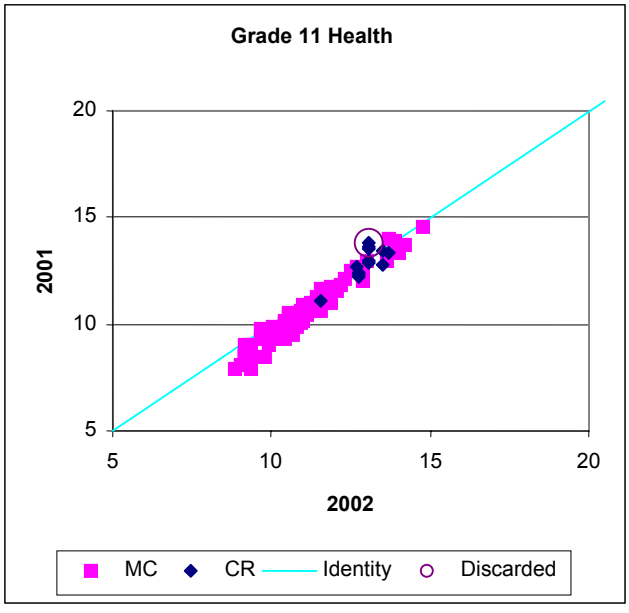

MEA Grade 4 Social Studies
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

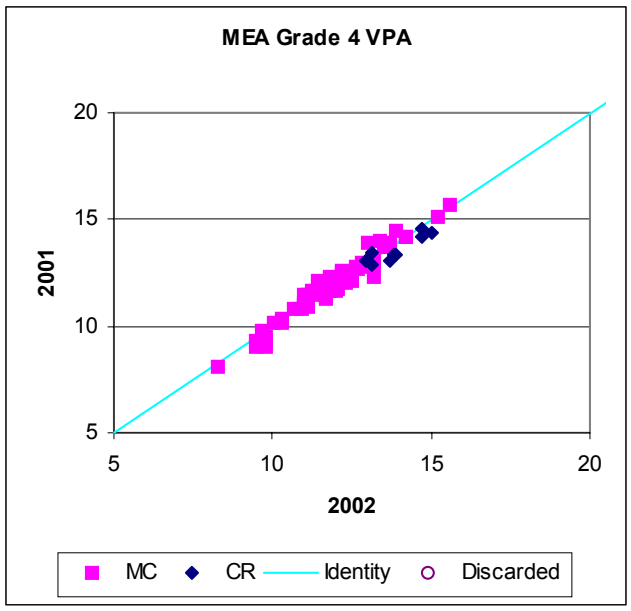

MEA Grade 8 Social Studies
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

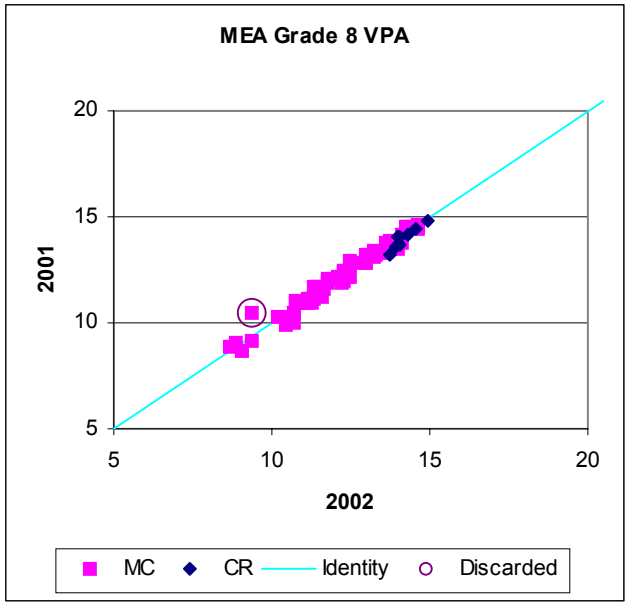

MEA Grade 11 Social Studies
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

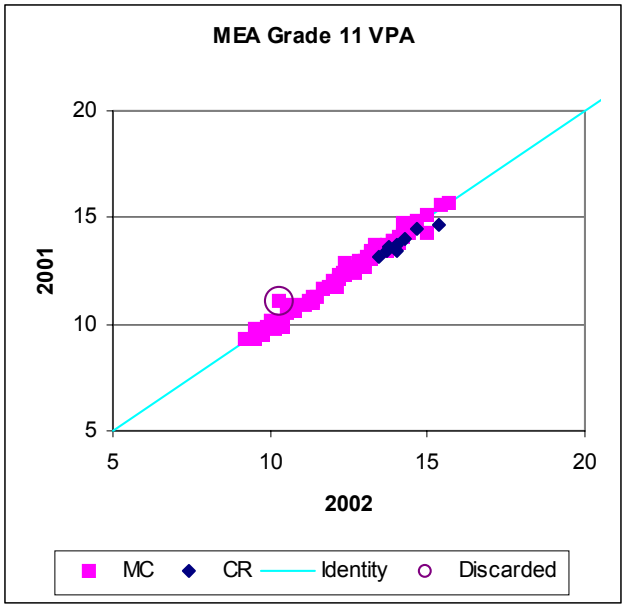

Grade 4 Health
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded


Grade 8 Health
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded


Grade 11 Health
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded


MEA Grade 4 VPA
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded


MEA Grade 8 VPA
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded


MEA Grade 11 VPA
20
15
10
5
2001
5
10
15
20
2002
MC
CR
Identity
Discarded

APPENDIX C

ACCURACY AND CONSISTENCY OF CLASSIFICATIONS

Accuracy and Consistency of Classifications

Grade 4 Reading

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.014023	0.00981	0.00000	0.000000	0.02383
Partially Meets the Standards	0.064774	0.53918	0.03731	0.000000	0.64124
Meets the Standards	0.000001	0.06049	0.22272	0.008947	0.29211
Exceeds the Standards	0.000000	0.00000	0.01076	0.031898	0.04266
	0.078797	0.60948	0.27079	0.040846	0.99984

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.01687	0.00674	6.78E-8	811E-23	0.02361
Partially Meets the Standards	0.07793	0.37052	0.06585	3.15E-8	0.51429
Meets the Standards	1.21E-6	0.04156	0.39311	0.00185	0.43652
Exceeds the Standards	485E-20	1.07E-7	0.019	0.00658	0.02558
Marginal	0.0948	0.41882	0.47795	0.00843	1

Accuracy	Cut #1	Cut #2	Cut #3
0.78707	0.91533	0.89258	0.97916

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.027424	0.05131	0.00007	0.000000	0.0788
Partially Meets the Standards	0.051308	0.49261	0.06555	0.000037	0.6095
Meets the Standards	0.000071	0.06555	0.19180	0.013338	0.2708
Exceeds the Standards	0.000000	0.00004	0.01334	0.027473	0.0408
	0.078803	0.60951	0.27076	0.040849	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.032990	0.03525	0.00013	8.6509E-12	0.06837
Partially Meets the Standards	0.061722	0.33844	0.11569	.000007650	0.51593
Meets the Standards	0.000085	0.04504	0.33856	.002752304	0.38646
Exceeds the Standards	0.000000	0.00003	0.02354	.005668640	0.02924
	0.094797	0.41876	0.47792	.008428594	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.71573	0.90281	0.83900	0.97367	0.52021

Accuracy and Consistency of Classifications

Grade 4 Writing

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.19604	0.06091	0.00008	1.0528E-11	0.25702
Partially Meets the Standards	0.11696	0.48883	0.07242	.000031605	0.67822
Meets the Standards	0.00006	0.02092	0.04333	.000358403	0.06467
Exceeds the Standards	0.00000	0.00000	0.00000	0	0.00000
	0.31306	0.57066	0.11582	.000390008	0.99991

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.1447	0.06718	0.00009	386E-13	0.21197
Partially Meets the Standards	0.08632	0.53916	0.08634	0.00012	0.71194
Meets the Standards	0.00004	0.02308	0.05166	0.00131	0.07609
Exceeds the Standards	0	0	0	0	0
Marginal	0.23106	0.62941	0.13809	0.00143	1

Accuracy	Cut #1	Cut #2	Cut #3
0.73552	0.84637	0.89033	0.99857

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.19092	0.11938	0.00274	.000000366	0.3131
Partially Meets the Standards	0.11938	0.38733	0.06387	.000089765	0.5707
Meets the Standards	0.00274	0.06387	0.04891	.000294030	0.1158
Exceeds the Standards	0.00000	0.00009	0.00029	.000005880	0.0004
	0.31304	0.57068	0.11582	.000390041	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.14090	0.13165	0.00327	.000001342	0.27587
Partially Meets the Standards	0.08812	0.42712	0.07616	.000329137	0.59179
Meets the Standards	0.00202	0.07045	0.05831	.001078129	0.13187
Exceeds the Standards	0.00000	0.00010	0.00035	.000021566	0.00047
	0.23104	0.62933	0.13808	.001430173	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.62644	0.77491	0.84767	0.99814	0.31528

Accuracy and Consistency of Classifications

Grade 4 Mathematics

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.27527	0.04459	0.00005	6.8434E-12	0.31989
Partially Meets the Standards	0.07433	0.34906	0.06441	.000033394	0.48785
Meets the Standards	0.00009	0.03901	0.14194	.007607460	0.18863
Exceeds the Standards	0.00000	0.00000	0.00164	.001852989	0.00350
	0.34968	0.43266	0.20804	.009493843	0.99986

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.22484	0.05064	0.00005	109E-13	0.27554
Partially Meets the Standards	0.06071	0.39643	0.06437	0.00005	0.52156
Meets the Standards	0.00007	0.0443	0.14185	0.0121	0.19831
Exceeds the Standards	127E-14	1.5E-6	0.00164	0.00295	0.00459
Marginal	0.28562	0.49137	0.20791	0.0151	1

Accuracy	Cut #1	Cut #2	Cut #3
0.76606	0.88852	0.89115	0.98621

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.26672	0.08096	0.00198	.000000451	0.3497
Partially Meets the Standards	0.08096	0.28131	0.07007	.000319958	0.4327
Meets the Standards	0.00198	0.07007	0.12909	.006916046	0.2081
Exceeds the Standards	0.00000	0.00032	0.00692	.002257347	0.0095
	0.34966	0.43266	0.20805	.009493802	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.21786	0.09193	0.00198	0.000001	0.31180
Partially Meets the Standards	0.06613	0.31946	0.07002	0.000509	0.45615
Meets the Standards	0.00161	0.07957	0.12897	0.010996	0.22119
Exceeds the Standards	0.00000	0.00036	0.00691	0.003589	0.01086
	0.28561	0.49133	0.20788	0.015095	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.66995	0.83833	0.84593	0.98122	0.48482

Accuracy and Consistency of Classifications

Grade 4 Science

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.26099	0.06778	0.000000	7.8686E-22	0.32874
Partially Meets the Standards	0.08359	0.54138	0.024853	.000000012	0.64990
Meets the Standards	0.00000	0.00710	0.014172	.000001892	0.02127
Exceeds the Standards	0.00000	0.00000	0.000000	0	0.00000
	0.34457	0.61626	0.039024	.000001904	0.99991

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.21168	0.07615	8.63E-8	183E-21	0.28783
Partially Meets the Standards	0.0678	0.60828	0.01762	2.89E-6	0.69371
Meets the Standards	3.91E-8	0.00798	0.01005	0.00044	0.01846
Exceeds the Standards	0	0	0	0	0
Marginal	0.27948	0.69241	0.02766	0.00044	1

Accuracy	Cut #1	Cut #2	Cut #3
0.83001	0.85605	0.97440	0.99956

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.23975	0.10478	0.000054	7.3019E-13	0.3446
Partially Meets the Standards	0.10478	0.48804	0.023529	.000000177	0.6164
Meets the Standards	0.00005	0.02353	0.015440	.000001725	0.0390
Exceeds the Standards	0.00000	0.00000	0.000002	.000000003	0.0000
	0.34458	0.61635	0.039025	.000001904	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.19443	0.11772	0.000038	1.6988E-10	0.31221
Partially Meets the Standards	0.08499	0.54822	0.016678	.000041083	0.64997
Meets the Standards	0.00004	0.02643	0.010944	.000401258	0.03782
Exceeds the Standards	0.00000	0.00000	0.000001	.000000778	0.00000
	0.27946	0.69237	0.027662	.000443119	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.75364	0.79720	0.95676	0.99956	0.46636

Accuracy and Consistency of Classifications

Grade 4 Social Studies

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.13947	0.02332	0.00000	9.1108E-14	0.16278
Partially Meets the Standards	0.09164	0.45459	0.05488	.000012353	0.60107
Meets the Standards	0.00007	0.06549	0.16040	.006046295	0.23203
Exceeds the Standards	0.00000	0.00000	0.00169	.002276897	0.00397
	0.23118	0.54340	0.21697	.008335546	0.99985

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.09503	0.02389	5.24E-6	112E-15	0.11892
Partially Meets the Standards	0.06244	0.46574	0.06969	0.00002	0.59789
Meets the Standards	0.00005	0.0671	0.2037	0.0074	0.27826
Exceeds the Standards	485E-16	1.03E-6	0.00214	0.00279	0.00493
Marginal	0.15752	0.55673	0.27555	0.01021	1

Accuracy	Cut #1	Cut #2	Cut #3
0.76725	0.91361	0.86313	0.99044

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.15353	0.07655	0.00111	.000000124	0.2312
Partially Meets the Standards	0.07655	0.38647	0.08012	.000267208	0.5435
Meets the Standards	0.00111	0.08012	0.12997	.005764008	0.2170
Exceeds the Standards	0.00000	0.00027	0.00576	.002305031	0.0083
	0.23119	0.54342	0.21697	.008336370	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.10460	0.07841	0.00141	0.000000	0.18444
Partially Meets the Standards	0.05215	0.39594	0.10175	0.000327	0.55019
Meets the Standards	0.00075	0.08209	0.16504	0.007055	0.25496
Exceeds the Standards	0.00000	0.00027	0.00732	0.002822	0.01042
	0.15751	0.55672	0.27551	0.010204	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.66843	0.86726	0.81338	0.98502	0.44207

Accuracy and Consistency of Classifications

Grade 4 Health

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.000000	0.00000	0.00000	0.000000	0.00000
Partially Meets the Standards	0.015560	0.60840	0.17456	0.000132	0.79871
Meets the Standards	0.000132	0.05148	0.12732	0.008049	0.18698
Exceeds the Standards	0.000000	0.00001	0.00437	0.009851	0.01423
	0.015692	0.65989	0.30625	0.018033	0.99992

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0	0	0	0	0
Partially Meets the Standards	0.01436	0.60932	0.17279	0.00016	0.79663
Meets the Standards	0.00012	0.05156	0.12604	0.00958	0.18731
Exceeds the Standards	588E-14	6.17E-6	0.00433	0.01173	0.01606
Marginal	0.01448	0.66089	0.30316	0.02147	1

Accuracy	Cut #1	Cut #2	Cut #3
0.74709	0.98552	0.77536	0.98593

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.000346	0.01225	0.00310	0.000002	0.0157
Partially Meets the Standards	0.012247	0.48639	0.16074	0.000546	0.6600
Meets the Standards	0.003095	0.16074	0.13370	0.008701	0.3063
Exceeds the Standards	0.000002	0.00055	0.00870	0.008783	0.0180
	0.015690	0.65992	0.30623	0.018033	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.000319	0.01226	0.00306	0.000002	0.01565
Partially Meets the Standards	0.011303	0.48706	0.15912	0.000650	0.65818
Meets the Standards	0.002857	0.16095	0.13235	0.010359	0.30655
Exceeds the Standards	0.000002	0.00055	0.00861	0.010454	0.01962
	0.014481	0.66082	0.30315	0.021465	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.63023	0.97051	0.67277	0.97982	0.21565

Accuracy and Consistency of Classifications

Grade 4 Visual and Performing Arts

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.18665	0.10408	0.00344	0.000070	||	0.29425
Partially Meets the Standards	0.11757	0.35767	0.08609	0.008978	||	0.57031
Meets the Standards	0.00372	0.06136	0.05523	0.015038	||	0.13535
Exceeds the Standards	0.00000	0.00000	0.00000	0.000000	||	0.00000
	0.30794	0.52310	0.14476	0.024085		0.99991

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.1849	0.09522	0.00435	0.0001	||	0.28456
Partially Meets the Standards	0.11646	0.32724	0.10869	0.01253	||	0.56492
Meets the Standards	0.00369	0.05613	0.06972	0.02098	||	0.15052
Exceeds the Standards	0	0	0	0	||	0
Marginal	0.30504	0.47859	0.18276	0.0336	||	1

Accuracy	Cut #1	Cut #2	Cut #3
0.58186	0.78019	0.81452	0.96640

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.15515	0.13556	0.01585	0.001405	||	0.3080
Partially Meets the Standards	0.13556	0.29535	0.08025	0.011982	||	0.5232
Meets the Standards	0.01585	0.08025	0.04017	0.008492	||	0.1448
Exceeds the Standards	0.00141	0.01198	0.00849	0.002206	||	0.0241
	0.30797	0.52314	0.14476	0.024085		1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	||	Marginal
Does Not Meet the Standards	0.15366	0.12401	0.02000-01	0.001961	||	0.29967
Partially Meets the Standards	0.13425	0.27020	0.10130	0.016716	||	0.52250
Meets the Standards	0.01570	0.07341	0.05071	0.011847	||	0.15168
Exceeds the Standards	0.00139	0.01096	0.01072	0.003078	||	0.02615
	0.30500	0.47858	0.18275	0.033601		1.00000

Consistency	Cut #1	Cut #2	Cut #3	||	kappa
0.47768	0.70265	0.75852	0.94640	||	0.17082

Accuracy and Consistency of Classifications

Grade 8 Reading

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.05428	0.02839	0.00000	0.000000	0.08267
Partially Meets the Standards	0.06499	0.49298	0.05765	0.000004	0.61560
Meets the Standards	0.00001	0.05377	0.20432	0.010357	0.26843
Exceeds the Standards	0.00000	0.00000	0.00882	0.024384	0.03320
	0.11928	0.57515	0.27078	0.034744	0.99990

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.05711	0.02173	4.07E-6	102E-17	0.07885
Partially Meets the Standards	0.06837	0.37737	0.08926	1.64E-6	0.53501
Meets the Standards	0.00001	0.04116	0.31636	0.00446	0.36199
Exceeds the Standards	159E-17	7.6E-7	0.01365	0.0105	0.02415
Marginal	0.12549	0.44026	0.41928	0.01496	1

Accuracy	Cut #1	Cut #2	Cut #3
0.76134	0.90988	0.86956	0.98189

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.05429	0.06454	0.00045	0.000000	0.1193
Partially Meets the Standards	0.06454	0.43323	0.07724	0.000140	0.5752
Meets the Standards	0.00045	0.07724	0.17975	0.013330	0.2708
Exceeds the Standards	0.00000	0.00014	0.01333	0.021271	0.0347
	0.11928	0.57515	0.27077	0.034741	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.05711	0.04941	0.00069	0.000000	0.10722
Partially Meets the Standards	0.06790	0.33160	0.11960	0.000060	0.51919
Meets the Standards	0.00047	0.05912	0.27832	0.005741	0.34368
Exceeds the Standards	0.00000	0.00011	0.02064	0.009161	0.02991
	0.12548	0.44024	0.41925	0.014962	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.67623	0.88151	0.81993	0.97345	0.47218

Accuracy and Consistency of Classifications

Grade 8 Writing

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.00000	0.00000	0.00000	0.000000	0.00000
Partially Meets the Standards	0.13870	0.52197	0.08720	0.000327	0.74829
Meets the Standards	0.00045	0.06008	0.16443	0.018158	0.24313
Exceeds the Standards	0.00000	0.00003	0.00404	0.004471	0.00854
	0.13915	0.58208	0.25567	0.022955	0.99996

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0	0	0	0	0
Partially Meets the Standards	0.10893	0.44408	0.13226	0.00011	0.68538
Meets the Standards	0.00035	0.05111	0.24941	0.0061	0.30697
Exceeds the Standards	724E-12	0.00002	0.00612	0.0015	0.00765
Marginal	0.10928	0.49521	0.38779	0.00771	1

Accuracy	Cut #1	Cut #2	Cut #3
0.69499	0.89072	0.81614	0.98764

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.03962	0.09355	0.00598	0.000013	0.1392
Partially Meets the Standards	0.09355	0.38971	0.09705	0.001819	0.5822
Meets the Standards	0.00598	0.09705	0.13739	0.015276	0.2557
Exceeds the Standards	0.00001	0.00182	0.01528	0.005848	0.0230
	0.13916	0.58213	0.25569	0.022956	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.03111	0.07957	0.00907	.000004441	0.11977
Partially Meets the Standards	0.07346	0.33148	0.14716	.000611186	0.55278
Meets the Standards	0.00470	0.08255	0.20837	.005130768	0.30076
Exceeds the Standards	0.00001	0.00155	0.02317	.001964092	0.02669
	0.10927	0.49515	0.38777	.007710488	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.57298	0.83317	0.75433	0.96953	0.28392

Accuracy and Consistency of Classifications

Grade 8 Mathematics

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.36591	0.03840	0.00001	1.005E-13	0.40430
Partially Meets the Standards	0.06097	0.31476	0.03799	.000007743	0.41376
Meets the Standards	0.00003	0.04253	0.13397	.005357742	0.18188
Exceeds the Standards	0.00000	0.00000	0.00000	0	0.00000
	0.42691	0.39568	0.17197	.005365485	0.99994

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.34135	0.03769	7.86E-6	175E-15	0.37905
Partially Meets the Standards	0.05688	0.30899	0.04505	0.00001	0.41093
Meets the Standards	0.00003	0.04174	0.1589	0.00935	0.21001
Exceeds the Standards	0	0	0	0	0
Marginal	0.39826	0.38842	0.20396	0.00936	1

Accuracy	Cut #1	Cut #2	Cut #3
0.80923	0.90539	0.91315	0.99064

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.35773	0.06857	0.00061	.000000050	0.4269
Partially Meets the Standards	0.06857	0.27203	0.05486	.000203282	0.3957
Meets the Standards	0.00061	0.05486	0.11201	.004471779	0.1720
Exceeds the Standards	0.00000	0.00020	0.00447	.000690222	0.0054
	0.42691	0.39567	0.17196	.005365332	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.33368	0.06731	0.00072	.000000087	0.40175
Partially Meets the Standards	0.06396	0.26703	0.06506	.000354588	0.39645
Meets the Standards	0.00057	0.05386	0.13284	.007801056	0.19509
Exceeds the Standards	0.00000	0.00020	0.00530	.001204014	0.00671
	0.39821	0.38839	0.20393	.009359745	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.73483	0.86742	0.87922	0.98634	0.58962

Accuracy and Consistency of Classifications

Grade 8 Science

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.24539	0.05799	0.00001	3.525E-14	0.30334
Partially Meets the Standards	0.07811	0.47913	0.04778	.000005266	0.60498
Meets the Standards	0.00001	0.02596	0.06363	.001275063	0.09087
Exceeds the Standards	0.00000	0.00000	0.00027	.000377655	0.00065
	0.32351	0.56307	0.11168	.001657983	0.99984

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.22169	0.06125	8.19E-6	479E-16	0.28295
Partially Meets the Standards	0.07058	0.50609	0.04736	7.16E-6	0.62403
Meets the Standards	8.76E-6	0.02742	0.06308	0.00173	0.09223
Exceeds the Standards	52E-17	7.46E-8	0.00027	0.00051	0.00078
Marginal	0.29228	0.59476	0.11071	0.00225	1

Accuracy	Cut #1	Cut #2	Cut #3
0.79138	0.86816	0.92520	0.99799

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.22870	0.09418	0.00062	.000000014	0.3235
Partially Meets the Standards	0.09418	0.41705	0.05182	.000057928	0.5631
Meets the Standards	0.00062	0.05182	0.05804	.001200676	0.1117
Exceeds the Standards	0.00000	0.00006	0.00120	.000399411	0.0017
	0.32350	0.56311	0.11168	.001658029	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.20660	0.09947	0.00061	.000000020	0.30672
Partially Meets the Standards	0.08508	0.44043	0.05137	.000078723	0.57702
Meets the Standards	0.00056	0.05473	0.05753	.001631737	0.11446
Exceeds the Standards	0.00000	0.00006	0.00119	.000542760	0.00179
	0.29225	0.59470	0.11071	.002253239	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.70519	0.81425	0.89258	0.99704	0.46831

Accuracy and Consistency of Classifications

Grade 8 Social Studies

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.23386	0.04073	0.00000	1.2807E-14	0.27460
Partially Meets the Standards	0.08304	0.43640	0.04414	.000004540	0.56360
Meets the Standards	0.00002	0.03954	0.11125	.004208565	0.15503
Exceeds the Standards	0.00000	0.00000	0.00222	.004514694	0.00674
	0.31691	0.51668	0.15762	.008727799	0.99997

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.20532	0.04189	4.53E-6	183E-16	0.24722
Partially Meets the Standards	0.07291	0.44878	0.04983	6.48E-6	0.57153
Meets the Standards	0.00002	0.04067	0.1256	0.00601	0.17229
Exceeds the Standards	521E-17	5.61E-7	0.00251	0.00645	0.00896
Marginal	0.27825	0.53134	0.17794	0.01246	1

Accuracy	Cut #1	Cut #2	Cut #3
0.78615	0.88517	0.90948	0.99147

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.23145	0.08487	0.00060	.000000019	0.3169
Partially Meets the Standards	0.08487	0.37415	0.05751	.000109300	0.5167
Meets the Standards	0.00060	0.05751	0.09486	.004631996	0.1576
Exceeds the Standards	0.00000	0.00011	0.00463	.003985405	0.0087
	0.31692	0.51663	0.15761	.008726720	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.20319	0.08728	0.00068	0.000000	0.29117
Partially Meets the Standards	0.07451	0.38477	0.06493	0.000156	0.52440
Meets the Standards	0.00053	0.05914	0.10710	0.006616	0.17339
Exceeds the Standards	0.00000	0.00011	0.00523	0.005692	0.01103
	0.27823	0.53130	0.17794	0.012464	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.70080	0.83698	0.87445	0.98789	0.50898

Accuracy and Consistency of Classifications

Grade 8 Health

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.000000	0.00000	0.00000	0	0.00000
Partially Meets the Standards	0.019325	0.59241	0.09358	.000020821	0.70532
Meets the Standards	0.000069	0.11278	0.17981	.001962662	0.29462
Exceeds the Standards	0.000000	0.00000	0.00000	0	0.00000
	0.019394	0.70518	0.27339	.001983482	0.99994

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0	0	0	0	0
Partially Meets the Standards	0.02482	0.51163	0.12441	0.00003	0.66089
Meets the Standards	0.00009	0.0974	0.23904	0.00258	0.33911
Exceeds the Standards	0	0	0	0	0
Marginal	0.02491	0.60903	0.36345	0.0026	1

Accuracy	Cut #1	Cut #2	Cut #3
0.75067	0.97509	0.77807	0.99740

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.001272	0.01723	0.00089	.000000104	0.0194
Partially Meets the Standards	0.017231	0.55383	0.13382	.000278592	0.7052
Meets the Standards	0.000892	0.13382	0.13705	.001623154	0.2734
Exceeds the Standards	0.000000	0.00028	0.00162	.000081688	0.0020
	0.019395	0.70516	0.27339	.001983538	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.001634	0.01488	0.00119	.000000137	0.01770
Partially Meets the Standards	0.022129	0.47833	0.17789	.000365734	0.67874
Meets the Standards	0.001146	0.11557	0.18219	.002130508	0.30105
Exceeds the Standards	0.000000	0.00024	0.00216	.000107229	0.00251
	0.024909	0.60902	0.36342	.002603608	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.66228	0.96065	0.70358	0.99510	0.29163

Accuracy and Consistency of Classifications

Grade 8 Visual and Performing Arts

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.20847	0.05501	0.00529	0.000120	0.26886
Partially Meets the Standards	0.16849	0.25214	0.10492	0.011023	0.53662
Meets the Standards	0.00914	0.05708	0.08521	0.030724	0.18216
Exceeds the Standards	0.00001	0.00045	0.00393	0.007900	0.01229
	0.38611	0.36467	0.19934	0.049766	0.99994

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.17366	0.06313	0.00592	0.00009	0.24281
Partially Meets the Standards	0.14036	0.28938	0.11751	0.0081	0.55534
Meets the Standards	0.00762	0.0655	0.09544	0.02257	0.19112
Exceeds the Standards	9.13E-6	0.00052	0.0044	0.0058	0.01073
Marginal	0.32165	0.41853	0.22327	0.03655	1

Accuracy	Cut #1	Cut #2	Cut #3
0.56428	0.78287	0.79474	0.96432

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.22937	0.11945	0.03385	0.003449	0.3861
Partially Meets the Standards	0.11945	0.15359	0.07808	0.013561	0.3647
Meets the Standards	0.03385	0.07808	0.06717	0.020237	0.1994
Exceeds the Standards	0.00345	0.01356	0.02024	0.012518	0.0498
	0.38612	0.36468	0.19934	0.049766	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.19104	0.13705	0.03792	0.002533	0.36859
Partially Meets the Standards	0.09949	0.17627	0.08745	0.009960	0.37320
Meets the Standards	0.02820	0.08960	0.07523	0.014864	0.20791
Exceeds the Standards	0.00287	0.01556	0.02267	0.009193	0.05030
	0.32160	0.41849	0.22326	0.036551	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.45177	0.69189	0.72588	0.93154	0.19020

Accuracy and Consistency of Classifications

Grade 11 Reading

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.002950	0.00187	0.00000	0.000000	0.00482
Partially Meets the Standards	0.051933	0.52136	0.04611	0.000001	0.61951
Meets the Standards	0.000008	0.08000	0.23862	0.010761	0.32941
Exceeds the Standards	0.000000	0.00000	0.01080	0.035431	0.04623
	0.054891	0.60323	0.29553	0.046193	0.99997

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.00424	0.00121	6.81E-8	248E-20	0.00545
Partially Meets the Standards	0.07463	0.33811	0.07928	5.41E-7	0.49201
Meets the Standards	0.00001	0.05187	0.41021	0.0051	0.46719
Exceeds the Standards	12E-17	5.89E-7	0.01856	0.01678	0.03534
Marginal	0.07888	0.3912	0.50805	0.02188	1

Accuracy	Cut #1	Cut #2	Cut #3
0.76934	0.92415	0.86884	0.97634

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.015491	0.03918	0.00022	0.000000	0.0549
Partially Meets the Standards	0.039185	0.48120	0.08279	0.000121	0.6033
Meets the Standards	0.000219	0.08279	0.19769	0.014822	0.2956
Exceeds the Standards	0.000000	0.00012	0.01482	0.031250	0.0462
	0.054895	0.60330	0.29553	0.046193	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.022259	0.02541	0.00038	0.000000	0.04804
Partially Meets the Standards	0.056297	0.31201	0.14233	0.000057	0.51072
Meets the Standards	0.000314	0.05368	0.33984	0.007018	0.40088
Exceeds the Standards	0.000000	0.00008	0.02548	0.014799	0.04036
	0.078870	0.39118	0.50803	0.021874	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.68895	0.91760	0.80315	0.96737	0.47446

Accuracy and Consistency of Classifications

Grade 11 Writing

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.05772	0.04411	0.00019	0.000000	0.10204
Partially Meets the Standards	0.06276	0.43561	0.08435	0.000638	0.58337
Meets the Standards	0.00043	0.08908	0.19287	0.023571	0.30597
Exceeds the Standards	0.00000	0.00006	0.00385	0.004676	0.00858
	0.12092	0.56886	0.28127	0.028884	0.99996

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.06954	0.03756	0.00023	2.21E-8	0.10734
Partially Meets the Standards	0.07561	0.37092	0.09956	0.00084	0.54694
Meets the Standards	0.00052	0.07585	0.22764	0.03098	0.33499
Exceeds the Standards	1.58E-9	0.00005	0.00455	0.00615	0.01074
Marginal	0.14567	0.48439	0.33198	0.03796	1

Accuracy	Cut #1	Cut #2	Cut #3
0.67425	0.88608	0.82295	0.96359

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.04727	0.07014	0.00346	0.000018	0.1209
Partially Meets the Standards	0.07014	0.38336	0.11176	0.003559	0.5689
Meets the Standards	0.00346	0.11176	0.14761	0.018425	0.2813
Exceeds the Standards	0.00002	0.00356	0.01842	0.006880	0.0289
	0.12090	0.56882	0.28125	0.028882	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.05695	0.05973	0.00408	0.000024	0.12080
Partially Meets the Standards	0.08452	0.32642	0.13190	0.004677	0.54756
Meets the Standards	0.00417	0.09515	0.17422	0.024216	0.29779
Exceeds the Standards	0.00002	0.00303	0.02175	0.009041	0.03384
	0.14566	0.48433	0.33195	0.037957	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.56670	0.84744	0.75692	0.94628	0.29776

Accuracy and Consistency of Classifications

Grade 11 Mathematics

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.36224	0.06067	0.00000	5.5695E-16	0.42291
Partially Meets the Standards	0.05702	0.33051	0.03817	.000001248	0.42572
Meets the Standards	0.00000	0.02551	0.11887	.004578590	0.14896
Exceeds the Standards	0.00000	0.00000	0.00124	.001081705	0.00232
	0.41927	0.41668	0.15828	.005661543	0.99991

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.33212	0.06259	4.65E-6	815E-18	0.39472
Partially Meets the Standards	0.05228	0.34103	0.04278	1.83E-6	0.43609
Meets the Standards	3.58E-6	0.02632	0.1332	0.0067	0.16622
Exceeds the Standards	92E-18	8.9E-8	0.00138	0.00158	0.00297
Marginal	0.3844	0.42994	0.17737	0.00829	1

Accuracy	Cut #1	Cut #2	Cut #3
0.80794	0.88512	0.93089	0.99191

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.33649	0.08246	0.00031	.000000002	0.4193
Partially Meets the Standards	0.08246	0.28961	0.04459	.000043690	0.4167
Meets the Standards	0.00031	0.04459	0.10904	.004332542	0.1583
Exceeds the Standards	0.00000	0.00004	0.00433	.001285315	0.0057
	0.41925	0.41671	0.15827	.005661549	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.30847	0.08508	0.00034	.000000003	0.39394
Partially Meets the Standards	0.07559	0.29877	0.04996	.000063941	0.42445
Meets the Standards	0.00028	0.04601	0.12219	.006340981	0.17483
Exceeds the Standards	0.00000	0.00005	0.00485	.001881123	0.00678
	0.38434	0.42990	0.17736	.008286047	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.73140	0.83869	0.90329	0.98869	0.57702

Accuracy and Consistency of Classifications

Grade 11 Science

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.29376	0.04565	0.000000	1.4286E-19	0.33942
Partially Meets the Standards	0.09816	0.46808	0.023666	.000000087	0.58984
Meets the Standards	0.00000	0.01991	0.048958	.000680447	0.06955
Exceeds the Standards	0.00000	0.00000	0.000380	.000679612	0.00106
	0.39192	0.53365	0.073004	.001360145	0.99987

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.22912	0.05166	1.8E-7	325E-21	0.28078
Partially Meets the Standards	0.07656	0.5296	0.02834	1.97E-7	0.6345
Meets the Standards	6.14E-7	0.02253	0.05864	0.00155	0.08272
Exceeds the Standards	753E-22	6.18E-9	0.00045	0.00155	0.002
Marginal	0.30568	0.60379	0.08743	0.00309	1

Accuracy	Cut #1	Cut #2	Cut #3
0.81890	0.87178	0.94912	0.99800

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.29486	0.09697	0.000103	3.9613E-11	0.3920
Partially Meets the Standards	0.09697	0.40625	0.030384	.000006040	0.5337
Meets the Standards	0.00010	0.03038	0.041740	.000774860	0.0730
Exceeds the Standards	0.00000	0.00001	0.000775	.000579238	0.0014
	0.39193	0.53361	0.073002	.001360138	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.22995	0.10971	0.000123	9.0068E-11	0.33981
Partially Meets the Standards	0.07562	0.45966	0.036385	.000013733	0.57172
Meets the Standards	0.00008	0.03437	0.049988	.001762152	0.08622
Exceeds the Standards	0.00000	0.00001	0.000928	.001317263	0.00225
	0.30565	0.60374	0.087423	.003093148	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.74097	0.81445	0.92901	0.99729	0.52331

Accuracy and Consistency of Classifications

Grade 11 Social Studies

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.27258	0.04407	0.00005	6.5912E-12	0.31671
Partially Meets the Standards	0.07048	0.33850	0.06535	.000035048	0.47437
Meets the Standards	0.00010	0.04485	0.15854	.005352020	0.20886
Exceeds the Standards	0.00000	0.00000	0.00000	0	0.00000
	0.34316	0.42742	0.22394	.005387068	0.99994

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.23619	0.04734	0.00005	929E-14	0.28358
Partially Meets the Standards	0.06107	0.36364	0.06884	0.00005	0.4936
Meets the Standards	0.00008	0.04819	0.16701	0.00754	0.22282
Exceeds the Standards	0	0	0	0	0
Marginal	0.29734	0.45916	0.2359	0.00759	1

Accuracy	Cut #1	Cut #2	Cut #3
0.76684	0.89146	0.88278	0.99241

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.26306	0.07817	0.00192	.000000490	0.3432
Partially Meets the Standards	0.07817	0.27429	0.07460	.000367939	0.4275
Meets the Standards	0.00192	0.07460	0.14282	.004591942	0.2240
Exceeds the Standards	0.00000	0.00037	0.00459	.000427008	0.0054
	0.34315	0.42743	0.22393	.005387379	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.22794	0.08397	0.00202	.000000690	0.31394
Partially Meets the Standards	0.06772	0.29462	0.07858	.000518680	0.44149
Meets the Standards	0.00166	0.08014	0.15045	.006472588	0.23873
Exceeds the Standards	0.00000	0.00040	0.00484	.000601888	0.00583
	0.29732	0.45912	0.23589	.007593845	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.67366	0.84460	0.83667	0.98778	0.49605

Accuracy and Consistency of Classifications

Grade 11 Health

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.000000	0.00000	0.00000	0	0.00000
Partially Meets the Standards	0.016491	0.56787	0.13455	.000033908	0.71899
Meets the Standards	0.000058	0.09296	0.18677	.001150131	0.28094
Exceeds the Standards	0.000000	0.00000	0.00000	0	0.00000
	0.016549	0.66083	0.32132	.001184039	0.99994

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0	0	0	0	0
Partially Meets the Standards	0.02861	0.56145	0.13161	0.0001	0.72178
Meets the Standards	0.0001	0.0919	0.18267	0.00354	0.27822
Exceeds the Standards	0	0	0	0	0
Marginal	0.02871	0.65336	0.31428	0.00365	1

Accuracy	Cut #1	Cut #2	Cut #3
0.74413	0.97129	0.77628	0.99635

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.000929	0.01446	0.00117	.000000133	0.0166
Partially Meets the Standards	0.014456	0.49292	0.15326	.000230283	0.6609
Meets the Standards	0.001166	0.15326	0.16592	.000939012	0.3213
Exceeds the Standards	0.000000	0.00023	0.00094	.000014678	0.0012
	0.016551	0.66087	0.32129	.001184104	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.001611	0.01429	0.00114	.000000409	0.01704
Partially Meets the Standards	0.025078	0.48730	0.14990	.000709414	0.66302
Meets the Standards	0.002022	0.15152	0.16229	.002892971	0.31875
Exceeds the Standards	0.000000	0.00023	0.00092	.000045210	0.00119
	0.028711	0.65334	0.31425	.003648004	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.65128	0.95747	0.69446	0.99525	0.25191

Accuracy and Consistency of Classifications

Grade 11 Visual and Performing Arts

Step 4

Predicted Classification X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.37225	0.23834	0.006857	0.000116	0.61755
Partially Meets the Standards	0.05869	0.15790	0.034302	0.003429	0.25433
Meets the Standards	0.00205	0.04401	0.057144	0.024853	0.12805
Exceeds the Standards	0.00000	0.00000	0.000000	0.000000	0.00000
	0.43300	0.44025	0.098303	0.028398	0.99994

Step 5

Actual Classification X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.35258	0.20403	0.01473	7.39E-6	0.57135
Partially Meets the Standards	0.05559	0.13516	0.0737	0.00022	0.26468
Meets the Standards	0.00194	0.03767	0.12278	0.00158	0.16398
Exceeds the Standards	0	0	0	0	0
Marginal	0.41012	0.37686	0.21121	0.00181	1

Accuracy	Cut #1	Cut #2	Cut #3
0.61052	0.72370	0.87173	0.99819

Step 6

X(1)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.24341	0.17865	0.010220	0.000711	0.4330
Partially Meets the Standards	0.17865	0.20932	0.043251	0.009033	0.4403
Meets the Standards	0.01022	0.04325	0.032455	0.012375	0.0983
Exceeds the Standards	0.00071	0.00903	0.012375	0.006280	0.0284
	0.43299	0.44025	0.098301	0.028399	1.0000

Step 7

X(0)

tstat	Does Not Meet the Standards	Partially Meets the Standards	Meets the Standards	Exceeds the Standards	Marginal
Does Not Meet the Standards	0.23053	0.15292	0.02195	.000045300	0.40549
Partially Meets the Standards	0.16919	0.17917	0.09293	.000575304	0.44189
Meets the Standards	0.00968	0.03702	0.06973	.000788093	0.11723
Exceeds the Standards	0.00067	0.00773	0.02659	.000399947	0.03540
	0.41007	0.37684	0.21120	.001808643	1.00000

Consistency	Cut #1	Cut #2	Cut #3	kappa
0.47987	0.64552	0.82938	0.96359	0.19027

APPENDIX D

DECISION RULES

Maine Educational Assessment

Decision Rules
Used for Reporting Results of the December 2001 Test Administration

Section I: General Rules

Issue	Rule #	Description	Pertinent Variables	Impact on Analyses/Aggregation	Impact on Parent Letter/Student Level Data	Impact on School/District Reports	Impact on Common Item Class Reports
Items							
Reading Items	1	Common	Form = 0 in IREF	Used to compute scaled scores, standard errors, and subscore information for individual students. Also used in computing subscores for school and district reports. Included in the criterion score for the item analyses.	Included in computing scaled scores, standard errors, and subscore information.	Included in computing subscores.	Item level scores of students in these items, and item difficulty summaries for class, school, district, and state are reported. Each multiple-choice item is reported as a "+" if correct and the response letter (A, B, C, or D) if incorrect. Blanks are reported as blanks and a multiple response is reported as "*". For constructed-response questions the number of points obtained is reported unless the students did not respond (reported as "B").
	2	Matrix	Form ne 0 in IREF	Used to compute subscores for school and district reports. Included in the criterion score for item analyses, except for common items. Some are equating items, and those items were used to equate scores from year to year.	None.	Included in computing subscores.	None.

Issue	Rule #	Description	Pertinent Variables	Impact on Analyses/Aggregation	Impact on Parent Letter/Student Level Data	Impact on School/District Reports	Impact on Common Item Class Reports
	3	Embedded FT	FT = "1" in IREF	Not used reporting MEA results. Included in the criterion score for item analyses for FT items only.	None.	Not included in computing subscores.	None.
Writing Items	4	Writing Prompt and Reading Item 33 (both common)	WP scores = Style1, style2	Used to compute scaled scores, standard errors, and subscore information for individual students. Also used in computing subscores for school and district reports.	Included in computing scaled scores, standard errors, and subscore information.	Included in computing subscores.	Item level scores of students in these items, and item difficulty summaries for class, school, district, and state are reported.
Health Education Items	5	Matrix	HEAmc01--HEAmc06, HEAor09, HEAor10	Used to compute subscores for school and district reports. Included in the criterion score for item analyses. Some are equating items, and those items were used to equate scores from year to year.	N/A	Included in computing subscores.	N/A
	6	Embedded FT	HEAmc07, HEAmc08, HAEmc11	Included in the criterion score for item analyses for FT items only.	N/A	Not included in computing subscores.	N/A
School Type							
Public Schools	7	Public schools that participated in the MEA (Public schools are required to participate in the MEA.)	Schstatus = "1"	Students from these schools are included in all state aggregation and all aggregation pertaining to the respective districts to which they belong unless otherwise dictated by other rules in this document.	Students in these schools will receive all information called for in the report unless otherwise dictated by other rules in this document.	Schools receive school reports unless otherwise dictated by other rules in this document. Data from these schools are used to compute district level data.	All pieces of information are provided.

Issue	Rule #	Description	Pertinent Variables	Impact on Analyses/Aggregation	Impact on Parent Letter/Student Level Data	Impact on School/District Reports	Impact on Common Item Class Reports
"Big 11" Schools	8	Private schools receiving state funding that participated in the MEA	Schstatus = "2"	Students from these schools are included in all state aggregation unless otherwise dictated by other rules in this document. Students in these schools are not included in any district level aggregation	All district level information will be blank (i.e., district scaled score average and district name).	Schools receive school reports unless otherwise dictated by other rules in this document, but students from these schools are not included in any district level aggregation.	Fields showing district level information will be blank (i.e., district level summaries and district name).
Private Schools	9	Private schools that participated in the MEA	Schstatus = "3"	Students from these schools will not be included in any district or state level aggregation except for the state level participation report.	All district level information will be blank (i.e., district scaled score average and district name).	Schools receive school reports unless otherwise dictated by other rules in this document, but students from these schools are not included in any district level aggregation.	Fields showing district level information will be blank (i.e., district level summaries and district name).
Exclusions							
Home Schooled	10	Home schooled students who participated in the MEA	Home = "1"	Home schooled students will not be included in any class, school, district, or state level aggregation except for the state level participation report.	Students will receive scaled scores and subscore information. There will be no school or district data, but there will be state data. In the school name field it should say "Home School" and in the district name field should be the name of the district of the school where the student took the test.	Students will not be included in any school or district level reports.	Each student will be reported in a separate class. There will be no class, school, or district level summaries. State level summaries will be provided. In the school name field it should say "Home School" and in the district name field should be the name of the district of the school where the student took the test. The class name field should be left blank. There will be an asterisk beside his/her name to indicate that he/she had been excluded in the computation for state level summaries.

Issue	Rule #	Description	Pertinent Variables	Impact on Analyses/Aggregation	Impact on Parent Letter/Student Level Data	Impact on School/District Reports	Impact on Common Item Class Reports
Did Not Participate (Specific to Content Area)	11	Student was marked as did not participate for a content area[2]	DNPrea = "1" DNPwri = "1" DNPhea = "1"	Student will not be included in any class, school, or district level aggregation of all types of scores or performance level for that content area.	Each student in this category will receive a report. Fields pertaining to that student's scores for the content area will be blank. In the performance level field it will say "Did Not Participate."	Student will not be included in any school or district level aggregation of any type of scores or performance level for that content area.	Each student in this category will be listed in the common item report under the class, school, and district indicated in his/her data. There will be an asterisk beside his/her name to indicate that he/she had been excluded in the computation for the class, school, district, and state level summaries. For these students the raw score field (i.e., "Points Earned") will say "DNP" and the scaled score and performance level fields will be blank.
Tested Incomplete (Specific to Content Area)	12	Student did not attempt at least one question in each session for a content area.[3]	TIrea = "1" TIwri = "1" TIhea = "1"	Student will not be included in any class, school, or district level aggregation of all types of scores or performance level for that content area.	Each student in this category will receive a report. Fields pertaining to that student's scores for the content area will be blank. In the performance level field it will say "Tested Incomplete."	Student will not be included in any school or district level aggregation of any type of scores or performance level for that content area.	Each student in this category will be listed in the common item report under the class, school, and district indicated in his/her data. There will be an asterisk beside his/her name to indicate that he/she had been excluded in the computation for the class, school, district, and state level summaries. For these students the raw score field (i.e., "Points Earned") will say "TI" and the scaled score and performance level fields will be blank.

[2] This does not include students who were marked as "Did Not Participate" but no content area is specified. Such students are not excluded unless otherwise dictated by other rules in this document. However, those students are counted as "Students who did not participate in all or part of the assessment due to [absence/other reasons]" in the participation report.

[3] For the writing assessment, each of the two prompts is considered a content area and a "Blank" flag is the indicator of not attempting. A student with a "Not Scorable" flag for a prompt is considered to have attempted that prompt and will not receive a "TI" exclusion based on that prompt.

Issue	Rule #	Description	Pertinent Variables	Impact on Analyses/Aggregation	Impact on Parent Letter/Student Level Data	Impact on School/District Reports	Impact on Common Item Class Reports
Home School **and** Did Not Participate (Specific to Content Area)	13	Home schooled student who was marked as did not participate for a content area	Home = "1" and DNPrea = "1" DNPwri = "1" DNPhea = "1"	These students will not be included in any class, school, district, or state level aggregation except for the state level participation report.	Each student in this category will receive a report. Fields pertaining to that student's scores for the content area will be blank. In the performance level field it will say "Did Not Participate." There will be no school or district data, but there will be state data. In the school name field it should say "Home School" and in the district name field should be the name of the district of the school where the student took the test.	Students will not be included in any school or district level reports.	Each student will be reported in a separate class. There will be no class, school, or district level summaries. State level summaries will be provided. In the school name field it should say "Home School" and in the district name field should be the name of the district of the school where the student took the test. The class name field should be left blank. There will be an asterisk beside his/her name to indicate that he/she had been excluded in the computation for state level summaries. For these students the raw score field (i.e., "Points Earned") will say "DNP" and the scaled score and performance level fields will be blank.
Home School **and** Tested Incomplete (Specific to Content Area)	14	Home schooled student who did not attempt at least one question in a each session for a content area	Home = "1" And TIrea = "1" TIwri = "1" TIhea = "1"	These students will not be included in any class, school, district, or state level aggregation except for the state level participation report.	Each student in this category will receive a report. Fields pertaining to that student's scores for the content area will be blank. In the performance level field it will say "Tested Incomplete." There will be no school or district data, but there will be state data. In the school name field it should say "Home School" and in the district name field	Students will not be included in any school or district level reports.	Each student will be reported in a separate class. There will be no class, school, or district level summaries. State level summaries will be provided. In the school name field it should say "Home School" and in the district name field should be the name of the district of the school where the student took the test. The class name field should be left blank. There will be an

Issue	Rule #	Description	Pertinent Variables	Impact on Analyses/Aggregation	Impact on Parent Letter/Student Level Data	Impact on School/District Reports	Impact on Common Item Class Reports
					should be the name of the district of the school where the student took the test.		asterisk beside his/her name to indicate that he/she had been excluded in the computation for state level summaries. For these students the raw score field (i.e., "Points Earned") will say "TI" and the scaled score and performance level fields will be blank.
Did not Participate **and** Tested Incomplete	15	Student who was marked as did not participate in a content area and did not attempt at lease one questions in each session in that content area	DNPrea = "1" and Tirea = "1" DNPwri = "1" and Tiwri = "1" DNPhea = "1" and Tihea = "1"	Student will not be included in any class, school, or district level aggregation of all types of scores or performance level for that content area.	Each student in this category will receive a report. Fields pertaining to that student's scores for the content area will be blank. In the performance level field it will say "Did Not Participate."	Student will not be included in any school or district level aggregation of any type of scores or performance level for that content area.	Each student in this category will be listed in the common item report under the class, school, and district indicated in his/her data. There will be an asterisk beside his/her name to indicate that he/she had been excluded in the computation for the class, school, district, and state level summaries. For these students the raw score field (i.e., "Points Earned") will say "DNP" and the scaled score and performance level fields will be blank.
Home School **and** Did Not Participate **and** Tested Incomplete	16	Home schooled student who was marked as did not participate in a content area and did not attempt at lease one questions in each session in that	Home = "1" and DNPrea = "1" and Tirea = "1" DNPwri = "1" and Tiwri = "1" DNPhea = "1"	These students will not be included in any class, school, district, or state level aggregation except for the state level participation report.	Each student in this category will receive a report. Fields pertaining to that student's scores for the content area will be blank. In the performance level field it will say "Did Not Participate." There will be no school or district data, but there will be state data. In the school name field it should say	Students will not be included in any school or district level reports.	Each student will be reported in a separate class. There will be no class, school, or district level summaries. State level summaries will be provided. In the school name field it should say "Home School" and in the district name field should be the name of the district of the school where the student took the test. The

Issue	Rule #	Description	Pertinent Variables	Impact on Analyses/Aggregation	Impact on Parent Letter/Student Level Data	Impact on School/District Reports	Impact on Common Item Class Reports
		content area	and Tihea = "1"		"Home School" and in the district name field should be the name of the district of the school where the student took the test.		class name field should be left blank. There will be an asterisk beside his/her name to indicate that he/she had been excluded in the computation for state level summaries. For these students the raw score field (i.e., "Points Earned") will say "DNP" and the scaled score and performance level fields will be blank.
Number of Students[4]							
Number of Students in School [District]	17	Less than five (<5)	Ntotal	No analysis will be performed except for item level summary data for the common item reports.	There will be no school [district] level data reported	No report will be generated for the school [district]	No impact.
		Five or more (≥5)	Ntotal	Other analyses will be performed depending on the number of **included** students in each content area. (See the two rules immediately below.)	Inclusion of school [district] level data will be dependent on the number **included** students, which is specific to content area. (See the two rules immediately below.)	A report will be generated for the school [district] and there will be data in the school [district] level participation summary. There might or might not be data that are content specific depending on the number of **included** students for the content area. (See the two rules immediately below.)	No impact.
Number of **Included** Students in School [District]	18	Less than five (<5)	Nincl	No content area specific aggregation will be performed except for item level summary data for the common item reports	There will be no school [district] level data reported for the content area	There will be no school [district] level data reported for the content area	No impact.

[4] Note that the rules on the (1) number of students, (2) number of included students, and (3) number of students in Reporting Category (or Questionnaire Items) are applied hierarchically. That is, the rule on the number of students in a reporting category is only relevant if there are five or more included students in the content area, and the rule on the number of included students on the content area is only relevant if there are five or more students in the school [district].

Issue	Rule #	Description	Pertinent Variables	Impact on Analyses/Aggregation	Impact on Parent Letter/Student Level Data	Impact on School/District Reports	Impact on Common Item Class Reports
(Specific to Content Area)		Five or more (≥5)	Nincl	All school [district] level aggregation of scaled scores and performance level will be done	School [district] level data will be reported for the content area	School [district] level data will be reported for the content area	No impact.
Number of Students in a **Reporting Category**[5]	19	Less than five (<5)		Count, scaled score, and performance level summaries **not** computed or reported for students in the category	N/A	The whole line for that category is left blank	N/A
		Five or more (≥5)		Count, scaled score, and performance level summaries computed and reported for students in the category	N/A	The whole line for that category is filled with the appropriate information	N/A
Number of Students in a **Questionnaire Response Category**	20	Less than five (<5)		No impact.	N/A	No impact.	N/A
		Five or more (≥5)		No impact.	N/A	No impact.	N/A

[5] Percentages across categories should sum up to 100% (withstanding rounding errors) except for categories under **Language minority/LEP students**, **Migrant**, and **Title I**.

Section II: Rules on Reporting Student Participation

Row Heading	Line #	Relation to Data	Determining Students Included in Each Category	Comments
Students enrolled on the first day of testing	1		All students in the grade level file.	• This is the only instance in the MEA reporting when students in private schools and home-schooled students are included in the state aggregate. • Should be equal to the sum of lines 14, 15, 21, 26, and 27.
Ethnicity	2		Sum of lines 3 through 10.	Should be equal to line 1.
White (non-Hispanic)	3	Ethnic = "1"		
Black (non-Hispanic)	4	Ethnic = "2"		
Hispanic	5	Ethnic = "3"		
Asian/Pacific Islander	6	Ethnic = "4"		
American Indian/Alaskan Native	7	Ethnic = "5"		
Multi-ethnic	8	Ethnic = "6"		
Other	9	Ethnic = "7"		
Not reported	10	Ethnic = " "		

Internet access at home	11		Sum of lines 12 and 13.	Should be equal to line 1.
Yes	12	Grades 4 and 8: StudQ02 = "A" Grade 11: StudQ05 = "A"		
No	13	Grades 4 and 8: StudQ02 ≠ "A" Grade 11: StudQ05 ≠ "A"		
Students who took all or part of the assessment without accommodation	14		Number of students minus the sum of lines 15, 21, 26, and 27.	
Students who took all or part of the assessment with accommodation	15	Accflag = "1"	Less than or equal to the sum of lines 16 through 20.	The categories under this heading are not mutually exclusive.
Identified disability (PET/IEP)	16	AccrPET = "1"		
LEP	17	AccrLEP = "1"		

504 plan	18	Accr504 =”1”		
Other	19	AccrOth = “1”		
Reason not reported	20	Accflag = “1” but blank for the four fields above		
Students recommended for participation in alternate assessment (PAAP)	21	Altflag = “1”	Less than or equal to the sum of lines 22 through 25.	The categories under this heading are not mutually exclusive.
Identified disability (PET/IEP)	22	AltrPET = “1”		
LEP	23	AltrLEP = “1”		
504 plan	24	Altr504 = “1”		
Reason not reported	25	Altflag = “1” but blank for the the three fields above		
Students who did not participate in all or part of the assessment due to absence	26	Ntabs = “1”		
Students who did not participate in all or part of the assessment due to other reasons	27	Ntoth = “1”		

APPENDIX E

QUALITY ASSURANCE CHECK LISTS

Parent Report Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

Review Steps – Complete each time the files are run	**Date(s) Step Completed**	**Comments**	**Date Prerelease Final Review Completed**
1. Compare the number of reports to the number of students in the file received from data processing. There should be one report for each student. (Check Decision Rules)			
2. Review/Proof the letter side of the report and compare to the shell			
3. On the letter side, check the bottom right corner box showing the State Summary Results. a. Match the percentages to the preliminary state numbers. b. Make sure the bar graph lines up with the scale. c. Review the placement of the bars and numbers to be sure everything is within the box and looks appropriate.			
4. Review/proof the performance side text and match to the shell.			

Parent Report Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth	**School** **Rev Date**	**School** **Rev Date**	**School** **Rev Date**	**School** **Rev Date**	**School** **Rev Date**
1. Match the following "Parent Report" elements to the corresponding elements in the "Common Item Class Report" for each content area. Note: The "Common Item Class Report" review should be completed before the "Parent Report" review. a. Student Name b. Student Grade c. School d. District e. Performance Level f. Scaled Score – three places 1. Numeric score 2. Visual - Diamond 3. Visual – bottom of bar in bar graph					
2. Verify that each student listed on the "Common Item Class Report" has a "Parent Report".					
3. For students who have TI or DNP on the "Common Item Class Report" a. Make sure that the student has no scaled score, performance level, diamonds and standard error bar, bar on bar graph or sub score diamonds. (Check Decision Rules for exceptions) b. Ensure the report has the correct notation. (Check Decision Rules)					
4. Match the School and District average scaled scores to the averages computed from the Common Item Report and the pre report calculation work.					
5. Match the State Average to the preliminary state numbers.					
6. Review the placement of the diamonds for the student scaled scores. Make sure they line up with the scale.					
7. Review height of the bar graphs for the average scaled scores. Make sure the height lines up with the scale.					
8. Using the standard error from the psychometrician, check the length					

Parent Report Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth	**School** **Rev Date**	**School** **Rev Date**	**School** **Rev Date**	**School** **Rev Date**	**School** **Rev Date**
and placement of the student standard error line through the student scaled score diamond.					
9. Using the sequel table or the recalculated sub categories for the student check the placement of the subcategory diamonds on the report.					
10. For Private Schools, there should be no district information. (Check Decision Rules)					
11. Schools with less than 5 students tested should have no School information. (Check Decision Rules)					
12. Districts with less than 5 students tested should have no District information. (Check Decision Rules)					
13. Home Schooled students should have student and state information only. School and District information should be blank. (Check Decision Rules)					

Label Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

General Label Quality Review Check List

Review Steps – Complete each time the file is run	**Date(s) Step Completed**	**Comments**	**Date Prerelease Final Review Completed**
1. Proof text and format of the label. Match to approved shell			
2. Make sure the same number of pages is in the file each time the file is run.			
3. Page through the file and check to see that each time the school name changes a new page is started.			

Label Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

Specific Label Quality Review Check List

General Review Steps – Complete steps in depth the first time the file is reviewed – thereafter, match initial review to new file to ensure correct data has not changed.	Sch Num Rev Date	Sch Num Rev Date	Sch Num Rev Date	Sch Num Rev Date	Sch Num Rev Date	Sch Num Rev Date	Sch Num Rev Date	Sch Num Rev Date	Sch Num Rev Date
1. Match the school and district information to the corresponding Common Item Report(s)									
2. Compare the number of Student Labels to the number of students listed on the Common Item Report(s) for the School. The numbers should match.									
3. Make sure the grade and the test administration date are correct.									
4. For each student, match the scaled scores and proficiency levels to the Common Item Report(s). They should be the same.									
5. Check to see that Home Schooled students appear at the end of the school and not in alphabetical order within the school.									

MEA Common Item Class Report Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

Review Steps – Complete each time the files are run.	Date(s) Step Completed	Comments	Date Prerelease Final Review Completed
1) Proof text and format of report, including legend if included in the file. Compare to shell.			
2) Compare "Content Standard & Performance Indicator" in the column heading area to I-Ref* or information supplied by the Program Manager			
3) Compare "Item Type" in the column heading area to I-Ref spreadsheet.			
4) Compare "Correct Multiple Choice Response" in the column heading area to I-Ref spreadsheet.			
5) Compare "Total Possible Points" in the column heading area to I-Ref spreadsheet.			
6) Compare total number of pages in the file to other Common Item files for the grade. All should have the same number of pages. The number of pages should not change from run to run unless students are added, assigned to different schools, or a school/district is added or deleted.			
7) Page through the PDF file and make sure the page numbers on the reports are sequential, e.g. 1 of 5, etc.			
8) Review the private schools and make sure there is no data in the district line.			
9) Check the State "Percent Correct/Avg. Score" to the State numbers computed on the spreadsheet. This information should not change from run to run as State data is frozen. Check with department manager if it does and document the reason.			

*I-Ref is the proprietary item bank relational database developed by Measured Progress.

MEA Common Item Class Report Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth	**Sch** (Num) **Class** (Name) **Rev Date**	**Sch** (Num) **Class** (Name) **Rev Date**	**Sch** (Num) **Class** (Name) **Rev Date**	**Sch** (Num) **Class** (Name) **Rev Date**	**Sch** (Num) **Class** (Name) **Rev Date**	**Sch** (Num) **Class** (Name) **Rev Date**	**Sch** (Num) **Class** (Name) **Rev Date**	**Sch** (Num) **Class** (Name) **Rev Date**	**Sch** (Num) **Class** (Name) **Rev Date**
1) Make sure the identifying information in the upper right hand box is complete. There should be information for Code, School, Date, Group Size, and Page. Only public schools will reflect a District Name and Class is an optional field.									
2) Check the home schooled list. If a student appears on the list for the school under review there should be a separate Common Item Report with the class indicated as "Home Schooled".									
3) Verify student names appear in alphabetical order in-groups of 5.									
4) Verify the Group Size by counting up the number of students.									
5) Verify the page numbers for each class and that all the pages are present.									
6) Highlight, with a yellow marker, each student listed on the exclusion list for the content area and class under review. If there is a "1" in the DNP column, there should be no "Points Earned", "Scaled Score", or "Performance Level" for these students. Instead, there should be a "DNP" in the "Points Earned" column.									
7) If a student's name appears on the exclusion list with a "1" in the "TI" column, there should be no "Points Earned", "Scaled Score", or "Performance Level" for this student. Instead, there should be a "TI" in the "Points Earned" column.									
8) Make sure each highlighted student has an "§" after his or her name.									
9) Count up the number of students that were highlighted									

MEA Common Item Class Report Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**
and subtract that number from the Group Size. Cross out the group size number and replace with the new number.									
10) Count up the number of pluses across the page for every third student starting randomly. Skip students whose rows you highlighted. Add the numerical scores to the total pluses and match your answer to the Points Earned for the student. It should be the same number you just calculated. If a student did not finish enough of the exam the "Points Earned" column will have a "TI".									
11) For each student in the above step, match the "Points Earned" to the conversion table to verify that the Scaled Score is correct. Then match the Scaled Score to the Performance level abbreviation to verify it is correct. ➢ 501 – 520 = D ➢ 521 – 540 = P ➢ 541 – 560 = M ➢ 561 – 580 = E									
12) For each "MC" "Item Type", review the "Item Number" column and make sure that no letters in the column match the "Correct MC Response". For example, if the "Correct MC Response" is "C" there should be no "C" in the column below.									
13) For each Item Type "SA" or "CR", review the numbers in the column and make sure none exceed the "Total Possible Points" for the column.									
14) Complete the appropriate attached form for ELA Writing by school. Note: All counts should exclude highlighted rows.									

MEA Common Item Class Report Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**
A) For each school – complete each step and indicate the results on the attached form.									
1) Indicate the number of students in each class adjusted for the exempt status (DNP and TI).									
2) Add the points in the "Writing Prompt" subcategory columns for the entire class									
3) Add the points in the "Extended Response" subcategory columns for the entire class									
4) Add the points in the "Total Writing" subcategory columns for the entire class									
5) Count up the total number of "E's", "M's", "P's", and "D's" in the "Performance Level Column"									
6) Add up the Scaled Scores for the entire class									
B) Total each column to get a school total.									
C) On one sheet, total all the schools to get a district total.									
D) Divide each column total by the Total Number of Students (Total Minus Highlighted) on both the school and district level to get the percent or average score.									
E) Match the Class, School and District percents or averages to the Common Item question on the report to verify the report is correct.									
15) Complete the appropriate attached form for Reading, Math, Science and Social Studies. Note: All counts should exclude highlighted rows.									
A) For each school – complete one form for each content area. Complete each step and indicate the results on the attached form.									

MEA Common Item Class Report Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**	**Sch (Num) Class (Name)** **Rev Date**
1) Indicate the number of students in the class adjusted for the exempt students.									
2) Pick question columns as indicated on the form. Be sure to pick the same columns for each class and school in a district. Vary the start column number by district so that all of the columns are chosen in the sample.									
3) For each column chosen count the number of "+'s" or add up the number of points depending on the column									
4) Count up the total number of "E's", "M's", "P's", and "D's" in the "Performance Level Column"									
5) Add up the Scaled Scores for the entire class									
B) Total the classes for each column to get a school total.									
C) On one sheet, total all the schools to get a district total.									
D) Divide each column total by the Total Number of Students (Total Minus Highlighted) on both the school and district level to get the percent or average score.									
E) Match the Class, School and District percents or averages each the Common Item question to the report to verify the report is correct.									
F) Match all numbers computed to the spreadsheet numbers.									
G) Match all summary and individual student results to the spreadsheet.									

MEA Common Item Class Report Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

ELA Writing

School_____
Name/Number

Class (Name)	Num. Stud. (Total Minus Highlighted)	Writing Prompt Total Points			Extended Response Question Total Points			Total Writing			Total Num. of "E's"	Total Num. of "M's"	Total Num. of "P's"	Total Num. of "D's"	Scaled Score Total
		Stylistic & Rhetorical	Standard English	Total	Stylistic & Rhetorical	Standard English	Total	Stylistic & Rhetorical	Standard English	Total					
School Tot / Percent															
District Tot / Percent															

MEA Common Item Class Report Quality Assurance Check List

District _____ Grade ____ Reviewer _____ Date _____

ELA Reading, Math, Science, or Social Studies

School _____
Name/Number

Class (Name)	Number of Students (Total Minus Highlighted	Pick an item number between 1 and 4 on the Common Item Report. Write the number in the first column below. Then write the number of every 3rd question after that. Add up the total number of "+" in each under the item number on the Common Item Report and post it below.										Total Number of "E's"	Total Number of "M's"	Total Number of "P's"	Total Number of "D's"	Scaled Score Total
School Total / Percent																
District Total / Percent																

MEA Reading and Writing School and District Reports Quality Assurance Checklist

District: _____ Grade: _____ Reviewer: _____ Date: _____

General Checklist

Review Steps – Complete each time the file is run	**Date(s) Step Completed**	**Comments**	**Date Prerelease Final Review Completed**
1. Check the number of pages in the file. It should not change from run to run.			
2. Proofread and match to approved DOE shell – Only do an in depth review on the first run - scan subsequent runs for obvious formatting issues.			
3. Scan through the file and check the page numbering. This is head to head duplex printed. Make sure the numbers are in the correct place and there is a blank page between reports if necessary.			
4. Make sure the Test Date on Page 1 is correct.			
5. Make sure there are no grid lines on the bar graphs on page 2.			

MEA Reading and Writing School and District Reports Quality Assurance Checklist

District: _____ Grade: _____ Reviewer: _____ Date: _____

Specific Checklist

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth – Note: On pages where there is school, district and State information all columns will be filled in on the School Report. On the District Report the School columns are blank. The plan will refer to these pages in steps as School/District data. On the School Report there are pages with school and State data. On the District Report there are pages with district and State data. The plan will refer to this in steps as School or District data.	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Dist** **Rev Date**
1. Using the "Common Item Class Report" match the identifying information on the cover page – ID (School or District Number), School, District, Grade, and Test Date									
2. Verify that the same identifying information is in the box in the upper right hand corner of every page – School (only on School Report), District, Grade, and Date									
3. Page 2: Executive Summary of School, District, and State Scores Box									
a. Using prior year reports, verify that the School/District/State Average Performance Scores for prior years for all content areas match.									
b. Using the "Common Item Class Report" review worksheets, verify that the School/District average Scores are correct.									
c. Using the State average scores verify that the State scores are correct. Note: the State scores should be frozen and not change during the review.									
d. Compute the "Cum. Avg." for School/District/State for each content area by adding the three years of scores and dividing by 3 (straight average). Verify that the averages on the report are correct.									
4. Page 2: Review the Bar Graphs.									
a. For Reading and Writing compare the performance percents for the school and district to the									

MEA Reading and Writing School and District Reports Quality Assurance Checklist

District: _____ Grade: _____ Reviewer: _____ Date: _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth – Note: On pages where there is school, district and State information all columns will be filled in on the School Report. On the District Report the School columns are blank. The plan will refer to these pages in steps as School/District data. On the School Report there are pages with school and State data. On the District Report there are pages with district and State data. The plan will refer to this in steps as School or District data.	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Dist** **Rev Date**
performance percents computed from the data on the "Common Item" Reports.									
b. For Health Education and State data, check with the psychometrician for the numbers. Check the placement of the bar and make sure it matches the scale.									
5. Page 3: Summary of Student Participation									
a. For each category listed on the page at the state, school, and district level, refer to the pre reporting spreadsheet prepared in advance for the number and percent for each category at the school and district level. Verify that the number reflected in the report matches the number computed.									
6. Pages 4 & 6 Reading/Writing Results – Students at each performance level									
a. Using the worksheets prepared from the "Common Item Class Report," verify the number and percent for each performance level are correct for the current year for the State, School, and District.									
7. Pages 4, 6, & 8 Reading/Writing/Health Results – Students at each Performance Level									
a. Using the previously gathered historical data, verify the number and percent for each performance category for each year is correct for the School, District and State.									
b. Check the State percent for the current year for each performance category									

MEA Reading and Writing School and District Reports Quality Assurance Checklist

District: _____ Grade: _____ Reviewer: _____ Date: _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth – Note: On pages where there is school, district and State information all columns will be filled in on the School Report. On the District Report the School columns are blank. The plan will refer to these pages in steps as School/District data. On the School Report there are pages with school and State data. On the District Report there are pages with district and State data. The plan will refer to this in steps as School or District data.	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Dist** **Rev Date**
c. Calculate the cumulative average at each performance level. This is a straight average.									
d. Add the performance percents for each year for School, District and State. The total should fall between 99 and 101.									
e. Compare the performance percents for the current year to the bar graphs on page 2. They should be the same.									
The following steps will be completed based on time available. They will be completed for at least one multi school District.									
8. Pages 4, 6, & 8 Reading/Writing/Health Results – Learning Results Content Standards for School or District									
a. Calculate the number of points possible for each sub category. Verify that it is the same number shown in the "Number of Points Possible" column for each sub category.									
b. Using the converted file sample, add up the total number of correct answers for the category questions and divide by the number of students tested. This number should appear in the "N" column for school and district.									
c. Divide the average number correct by the total number of possible points for the category. Multiply the result by 100. This will give you the number in the "%" column for school/district.									
d. Using the spreadsheet, verify that the State									

MEA Reading and Writing School and District Reports Quality Assurance Checklist

District: _____ Grade: _____ Reviewer: _____ Date: _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth – Note: On pages where there is school, district and State information all columns will be filled in on the School Report. On the District Report the School columns are blank. The plan will refer to these pages in steps as School/District data. On the School Report there are pages with school and State data. On the District Report there are pages with district and State data. The plan will refer to this in steps as School or District data.	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Dist** **Rev Date**
Learning Results numbers are correct.									
9. Pages 5, 7, & 9 – Reporting Categories									
a. Using the spreadsheet results for the state, district and school verify as correct:									
1. The percent of students in the category									
2. The average scaled score									
3. The percent that meet or exceed the standard									
4. The percent that partially meets the standard									
5. The percent that does not meet the standard									
6. Based on the decision rules and calculation method, where possible, verify that the percentages add up to between 99 and 101.									
10. Pages 5, 7, & 9 – Questionnaire Items									
a. Using the spreadsheet for each content area and the calculation method specified in the decision rules, calculate the percent of students who chose each response to a question. Verify the state results.									

MEA Reading and Writing School and District Reports Quality Assurance Checklist

District: _____ Grade: _____ Reviewer: _____ Date: _____

Specific Checklist

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth – Note: On pages where there is school, district and State information all columns will be filled in on the School Report. On the District Report the School columns are blank. The plan will refer to these pages in steps as School/District data. On the School Report there are pages with school and State data. On the District Report there are pages with district and State data. The plan will refer to this in steps as School or District data.	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Dist** **Rev Date**
1. Using the "Common Item Class Report" match the identifying information on the cover page – ID (School or District Number), School, District, Grade, and Test Date									
2. Verify that the same identifying information is in the box in the upper right hand corner of every page – School (only on School Report), District, Grade, and Date									
3. Page 2: Executive Summary of School, District, and State Scores Box									
a. Using prior year reports, verify that the School/District/State Average Performance Scores for prior years for all content areas match.									
b. Using the "Common Item Class Report" review worksheets, verify that the School/District average Scores are correct.									
c. Using the State average scores verify that the State scores are correct. Note: the State scores should be frozen and not change during the review.									
d. Compute the "Cum. Avg." for School/District/State for each content area by adding the three years of scores and dividing by 3 (straight average). Verify that the averages on the report are correct.									
4. Page 2: Review the Bar Graphs.									
a. For Reading and Writing compare the performance percents for the school and district to the performance percents computed from the data on									

MEA Reading and Writing School and District Reports Quality Assurance Checklist

District: _____ Grade: _____ Reviewer: _____ Date: _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth – Note: On pages where there is school, district and State information all columns will be filled in on the School Report. On the District Report the School columns are blank. The plan will refer to these pages in steps as School/District data. On the School Report there are pages with school and State data. On the District Report there are pages with district and State data. The plan will refer to this in steps as School or District data.	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Dist** **Rev Date**
the "Common Item" Reports.									
b. For Health Education and State data, check with the psychometrician for the numbers. Check the placement of the bar and make sure it matches the scale.									
5. Page 3: Summary of Student Participation									
a. For each category listed on the page at the state, school, and district level, refer to the pre reporting spreadsheet prepared in advance for the number and percent for each category at the school and district level. Verify that the number reflected in the report matches the number computed.									
6. Pages 4 & 6 Reading/Writing Results – Students at each performance level									
a. Using the worksheets prepared from the "Common Item Class Report," verify the number and percent for each performance level are correct for the current year for the State, School, and District.									
7. Pages 4, 6, & 8 Reading/Writing/Health Results – Students at each Performance Level									
a. Using the previously gathered historical data, verify the number and percent for each performance category for each year is correct for the School, District and State.									
b. Check the State percent for the current year for each performance category									
c. Calculate the cumulative average at each									

MEA Reading and Writing School and District Reports Quality Assurance Checklist

District: _____ Grade: _____ Reviewer: _____ Date: _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth – Note: On pages where there is school, district and State information all columns will be filled in on the School Report. On the District Report the School columns are blank. The plan will refer to these pages in steps as School/District data. On the School Report there are pages with school and State data. On the District Report there are pages with district and State data. The plan will refer to this in steps as School or District data.	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Dist** **Rev Date**
performance level. This is a straight average.									
d. Add the performance percents for each year for School, District, and State. The total should fall between 99 and 101.									
e. Compare the performance percents for the current year to the bar graphs on page 2. They should be the same.									
The following steps will be completed based on time available. They will be completed for at least one multi school District.									
8. Pages 4, 6, & 8 Reading/Writing/Health Results – Learning Results Content Standards for School or District									
a. Calculate the number of points possible for each sub category. Verify that it is the same number shown in the "Number of Points Possible" column for each sub category.									
b. Using the converted file sample, add up the total number of correct answers for the category questions and divide by the number of students tested. This number should appear in the "N" column for school and district.									
c. Divide the average number correct by the total number of possible points for the category. Multiply the result by 100. This will give you the number in the "%" column for school/district.									
d. Using the spreadsheet, verify that the State Learning Results numbers are correct.									

MEA Reading and Writing School and District Reports Quality Assurance Checklist

District: _____ Grade: _____ Reviewer: _____ Date: _____

Specific Review Steps – Complete in depth the first time the file is run – thereafter match initial review data that should not have changed to the new file and check the corrected data in depth – Note: On pages where there is school, district and State information all columns will be filled in on the School Report. On the District Report the School columns are blank. The plan will refer to these pages in steps as School/District data. On the School Report there are pages with school and State data. On the District Report there are pages with district and State data. The plan will refer to this in steps as School or District data.	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Sch Num** **Rev Date**	**Dist** **Rev Date**
9. Pages 5, 7, & 9 – Reporting Categories									
a. Using the spreadsheet results for the state, district and school verify as correct:									
1. The percent of students in the category									
2. The average scaled score									
3. The percent that meet or exceed the standard									
4. The percent that partially meets the standard									
5. The percent that does not meet the standard									
6. Based on the decision rules and calculation method, where possible, verify that the percentages add up to between 99 and 101.									
10. Pages 5, 7, & 9 – Questionnaire Items									
a. Using the spreadsheet for each content area and the calculation method specified in the decision rules, calculate the percent of students who chose each response to a question. Verify the state results.									

APPENDIX F

STANDARD SETTING

STANDARD SETTING

The Maine Department of Education, in an 18-month process with extensive input from educators and policy makers throughout the state, created four performance levels to describe student achievement:

- Does Not Meet the Standards,
- Partially Meets the Standards,
- Meets the Standards, and
- Exceeds the Standards.

Four policy considerations the department set for performance standards were that they be

- concrete,
- consistent,
- challenging, and
- obtainable.

The process used to determine the MEA scores necessary for each performance level was developed with these policy considerations in mind. Two sources of data were gathered.

- Twenty-one panels consisting of about 300 educators, parents, businesspeople, and policy makers systematically looked at samples of student work and rated the work against the four Maine performance level descriptors.
- About 5,000 additional teachers rated student classroom work against those same performance level descriptors.

The results of these two approaches were averaged and then adjusted to minimize any inconsistency of the standards across the different grade levels. This last adjustment was accomplished by averaging the results for each grade with the results for the other two grades. The effect of this adjustment was kept small by counting the results of the grade under consideration four times as heavily as the results of either other grade.

PERFORMANCE LEVELS DEFINITIONS

The following charts contain the subject-specific performance level definitions.

CHART F-1
READING

Exceeds the Standards—The quality of a student's work at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The work demonstrates exemplary accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate. (Scaled scores: 561–580.)

Meets the Standards—The quality of a student's work at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The work demonstrates a consistent accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate. (Scaled scores: 541–560.)

Partially Meets the Standards—The quality of a student's work at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The work demonstrates inconsistent accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate. (Scaled scores: 521–540.)

Does Not Meet the Standards—The quality of a student's work at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in English language arts (reading). The work demonstrates limited accomplishment in the comprehension of literary and informational texts, in the use of the skills and strategies of reading to answer questions, and in the demonstration of understanding of how words and images communicate. (Scaled scores: 501-520.)

CHART F-2
WRITING

Exceeds the Standards—The quality of a student's written compositions at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates exemplary accomplishment in both the development of the topic/idea and the use of Standard English conventions in first-draft writing. (Scaled scores:561–580.)

Meets the Standards—The quality of a student's written compositions at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates proficiency in both the development of the topic/idea and the use of Standard English conventions in first-draft writing. (Scaled scores:541–560.)

Partially Meets the Standards—The quality of a student's written compositions at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates writing skills that may show moderate development of topic/ideas and/or some errors in Standard English conventions that may interfere with communication. (Scaled scores:521–540.)

Does Not Meet the Standards—The quality of a student's written compositions at this level does not meet the standards of performance as identified for Maine's *Learning Results* in English language arts (writing). The student's work demonstrates writing skills that show limited development of topic/idea and/or many errors in Standard English conventions that interfere with communication of ideas. (Scaled scores:501–520.)

CHART F-3
HEALTH EDUCATION

Exceeds the Standards—The quality of a student's work at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates exemplary knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis, and risk reduction. (Scaled scores:561–580.)

Meets the Standards—The quality of a student's work at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates consistent knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis, and risk reduction. (Scaled scores:541–560.)

Partially Meets the Standards—The quality of a student's work at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates partial and/or inconsistent knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis, and risk reduction. (Scaled scores:521–540.)

Does Not Meet the Standards—The quality of a student's work at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in health education. The student demonstrates a limited knowledge of content and skills related to health promotion and disease prevention including communication, decision making, analysis, and risk reduction. (Scaled scores:501–520.)

CHART F-4
MATHEMATICS

Exceeds the Standards—The quality of a student's work at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in mathematics. The student's overall performance demonstrates exemplary knowledge of content, process, problem-solving, and communication skills. (Scaled scores:561–580.)

Meets the Standards—The quality of a student's work at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in mathematics. The student's work consistently shows complete knowledge of mathematical content, process, reasoning, and communication skills, as well as problem-solving abilities. (Scaled scores:541–560.)

Partially Meets the Standards—The quality of a student's work at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in mathematics. The student's work demonstrates a partial and/or inconsistent knowledge of mathematical content, process, reasoning, and communication skills, and problem-solving abilities. (Scaled scores:521–540.)

Does Not Meet the Standards—The quality of a student's work at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in mathematics. The student's work demonstrates a limited knowledge of mathematical content, process, reasoning, and communication skills, as well as problem-solving ability. (Scaled scores:501–520.)

CHART F-5
SCIENCE & TECHNOLOGY

Exceeds the Standards—The quality of a student's work at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in science and technology. The student demonstrates exemplary knowledge of content including life, physical, and earth/space sciences and scientific inquiry, reasoning, and communication skills. (Scaled scores:561–580.)

Meets the Standards—The quality of a student's work at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in science and technology. The student demonstrates consistent knowledge of content including life, physical, and earth/space sciences and scientific inquiry, reasoning, and communication skills. (Scaled scores:541–560.)

Partially Meets the Standards—The quality of a student's work at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in science and technology. The student demonstrates partial and/or inconsistent knowledge of content including life, physical, and earth/space sciences and scientific inquiry, reasoning, and communication skills. (Scaled scores:521–540.)

Does Not Meet the Standards—The quality of a student's work at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in science and technology. The student demonstrates limited knowledge of content including life, physical, and earth/space sciences and scientific inquiry, reasoning, and communication skills. (Scaled scores:501–520.)

CHART F-6
SOCIAL STUDIES

Exceeds the Standards—The quality of a student's work at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in social studies. The student demonstrates exemplary knowledge of content of major social studies concepts, consistently applies complex thinking skills, and communicates ideas clearly in all situations. (Scaled scores:561–580.)

Meets the Standards—The quality of a student's work at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in social studies. The student demonstrates consistent knowledge of content of major social studies concepts, usually applies complex thinking skills, and communicates ideas clearly in most situations. (Scaled scores:541–560.)

Partially Meets the Standards—The quality of a student's work at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in social studies. The student demonstrates some knowledge of content of major social studies concepts, inconsistently applies complex thinking skills, and communicates ideas clearly in some situations. (Scaled scores:521–540.)

Does Not Meet the Standards—The quality of a student's work at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in social studies. The student demonstrates a limited knowledge of content of major social studies concepts, does not apply complex thinking skills, and communicates ideas clearly in few or no situations. (Scaled scores:501–520.)

CHART F-7
VISUAL & PERFORMING ARTS

Exceeds the Standards—The quality of a student's work at this level of proficiency exceeds the standards of performance as identified for Maine's *Learning Results* in visual and performing arts. The student demonstrates exemplary knowledge of content and application of skills of the visual and performing arts, including creative expression, cultural heritage, and criticism and aesthetics. (Scaled scores:561–580.)

Meets the Standards—The quality of a student's work at this level of proficiency meets the standards of performance as identified for Maine's *Learning Results* in visual and performing arts. The student demonstrates consistent knowledge of content and application of skills of the visual and performing arts, including creative expression, cultural heritage, and criticism and aesthetics. (Scaled scores:541–560.)

Partially Meets the Standards—The quality of a student's work at this level of proficiency partially meets the standards of performance as identified for Maine's *Learning Results* in visual and performing arts. The student demonstrates partial and/or inconsistent knowledge of content and application of skills of the visual and performing arts, including creative expression, cultural heritage, and criticism and aesthetics. (Scaled scores:521–540.)

Does Not Meet the Standards—The quality of a student's work at this level of proficiency does not meet the standards of performance as identified for Maine's *Learning Results* in visual and performing arts. The student demonstrates limited knowledge of content and application of skills of the visual and performing arts, including creative expression, cultural heritage, and criticism and aesthetics. (Scaled scores:501–520.)

STANDARD SETTING METHODS

There were two standard setting methods used for the MEA: the Body of Work (BoW) method (Kingston, Kahl, Sweeney, & Bay, 2000) and the Contrasting Group (CG) method (Livingston & Zieky, 1982). Threshold scores resulting from the two methods were aggregated to obtain the minimum scores for each performance level.

The two methods and their implementations are described in the following sections. The threshold scores that were recommended to and accepted by the DOE are also presented.

CONTRASTING GROUP (CG)

The contrasting group method is based on the notion that examinees can be divided into two contrasting groups (Livingston & Zieky, 1982). For example, for the MEA these two groups could be the group of examinees that meets the standards (this includes those who exceed the standards) and the group of students that does not (this includes those who partially meet the standards and those who do not meet the standards).

Prior to the implementation of the BoW standard setting method, student rosters were sent to select schools with a request for teachers to assign performance levels to selected students in different subject areas. The instructions given to the teachers were as follows:

1. Carefully review the Maine *Learning Results* for this content area.
2. Carefully review the performance level definitions.
3. For each student listed, indicate the performance level that matches the student's achievement of the Maine *Learning Results*. (1 = Exceeds the Standards; 2 = Meets the Standard; 3 = Partially Meets the Standard; 4 = Does Not Meet the Standard)
4. Return the completed form to your building principal.

Included in the instructions is the information that the task of assigning performance levels was to be performed by the teacher who is currently teaching or who most recently taught this content area to the identified student. Teachers and principals involved in this study were told that information collected would be used along with information collected during standard setting sessions on July 26-29, 1999, to establish the performance level cutscores for the MEA.

A total of 73 schools in Maine were selected and asked to participate in this study: 44 for grade 4, 12 for grade 8, and 17 for grade 11, across the six subject areas. The number of students selected for this study for each grade level and subject combination is presented in Table F-1. These are the numbers of students that teachers have to assign to different performance levels.

Data collected from this effort were analyzed to obtain threshold scores for each performance level in each grade and subject area. These thresholds were combined with thresholds resulting from the BoW method to obtain the final thresholds recommended to the DOE. The method of combining the thresholds is discussed later in this chapter.

Table F-1 Number of Selected Students for the Contrasting Group			
Subject	**Grade 4**	**Grade 8**	**Grade 11**
Reading	330	340	328
Mathematics	328	326	338
Science and Technology	314	333	330
Social Studies	315	330	330
Health Education	312	332	357
Visual and Performing Arts	310	379	381

BODY OF WORK (BOW)

On July 26-29, 1999, panels were assembled for the implementation of the Body of Work (BoW) standard-setting method. The hallmark of the BoW method is that panelists examine complete student response sets (student responses to multiple-choice questions and samples of actual student work on open-response questions) and match each student response set to one of the MEA performance level categories. This is done in three major steps: (1) training/calibration, (2) range finding, and (3) pinpointing.

TRAINING/CALIBRATION

During this first phase of the MEA standard-setting process, panelists reviewed all MEA test questions for their assigned content area and grade level, and content- and grade-specific descriptors for each performance level. Panelists were given the opportunity to discuss and comment on test questions and descriptors. Next, to ensure that panelists attained a common interpretation of performance descriptors and the relationship of those descriptors to student work, panel members individually assigned performance levels to a set of six sample student responses. Panelists then compared their individual results and discussed at length how the performance level descriptors supported their conclusions.

RANGE-FINDING

During the range-finding phase of standard setting, identical sets of student work that spanned the score continuum were provided to each panelist. Panelists were asked to independently categorize the sets as Exceeds the Standards, Meets the Standards, Partially Meets the Standards, or Does Not Meet the Standards, based on the performance level descriptors. This process revealed which levels of student work generated the most agreement and which generated the most disagreement among panelists. The results were documented, and the levels of the sets of work that generated the most disagreement defined the score intervals in which the threshold scores must fall.

PINPOINTING

Additional sets of student work from score ranges that generated disagreement were presented to panelists. Panelists assigned performance levels to these sets of responses. The minimum score for each performance level was precisely pinpointed by determining the score around which there was, collectively, the maximum disagreement between panelists. This is the point that best represents the transition from response sets at a higher level to those at a lower level.

PANELISTS

Twenty-one panels were convened to set performance standards for the MEA—one panel for each grade level (4, 8, and 11) in seven areas—(1) reading, (2) writing, (3) mathematics, (4) science, (5) social studies, (6) health, and (7) visual and performing arts. The panels were composed of educators, parents and business leaders, and members of the general public.

IMPLEMENTATION

Following is a detailed description of the steps followed in implementing the MEA BoW standard-setting design.

BEFORE THE MEETING

1. For each subject-grade combination (e.g., grade 8 mathematics) pinpointing folders were prepared from samples of student work. This sample was double-scored to increase the accuracy of the standard-setting process. Any

students whose body of work was of uneven quality (for example, some open-response questions with scores of four and others with scores of one) were excluded, as were students whose open-response and multiple-choice responses were particularly discrepant. Folders ranged in scores from the highest obtained score in the remaining sample to the "approximately chance level" (0.25 times the number of multiple-choice items plus one times the number of open-response items). Each folder consisted of five sets of student work at each of four score points (e.g., five 12s, five 13s, five 14s, and five 15s), with the exception of the top folder (folder with highest scores). The top folder differed because there often were fewer than five papers available at any particular score point. Thus, the twenty papers in the top folder covered a wider range of scores. Approximately ten pinpointing folders were created for each subject-grade combination.

2. Range-finding folders were prepared from the pinpointing folders. The highest-scoring and two lowest-scoring papers were selected from each pinpointing folder. Thus, range-finding folders had about thirty samples of student work in each.
3. For each subject-grade combination, six student response sets spanning the range of performance were identified from the pinpointing folders. The facilitator reviewed the sets and prepared training notes consisting of points to be made during discussion of those student response sets. Focus was on ways responses illustrate characteristics described in the performance level definitions.
4. The Maine Department of Education created a list of members of each panel (one panel per subject area, four subject areas per grade, and three grades), ensuring each group had the proper diversity of membership (educator, parent, policy-maker, businessperson, ethnicity, gender, etc.). Color-coded name tags were provided to panel members.

GENERAL MEETING

Before the panels broke into separate groups, there was a general session at which logistical issues were addressed and the standard-setting procedures explained by the chief of standard setting. Major steps of the panel meeting portion of the meeting were described.

PANEL MEETING

1. Facilitators distributed the descriptor of a four-point response to each open-response question. Panel members were asked to review and discuss the test questions—open-response and multiple-choice. (Panelists had been asked to answer the questions before the meeting, and they were to have brought with them the tests and the performance level definitions. Additional copies were distributed to those who needed them.)
2. The facilitators led a discussion of the performance level definitions.
3. Training folders were distributed to every judge. The multiple-choice display at the end of a set was pointed out. Facilitators explained that it too should be considered when judgments are being made about the student work.
4. Judges were asked to rank independently the six previously identified student response sets based on overall quality, keeping in mind the performance level descriptions. Each judge listed the six student serial numbers in rank order from high to low performance on a separate piece of paper.
5. While the judges rank ordered the six student response sets, the facilitator wrote the serial numbers of the six sets on an overhead transparency in a vertical list in order from highest performance to lowest performance. When the judges completed their rankings, the facilitators showed the score rankings on the overhead projector and had the judges note the extent of agreement.
6. Judges were asked to assign each of the six response sets to a performance level. They each wrote the performance level initials (E, M, P, or D) next to the student serial numbers they listed in rank order in step 4.
7. Facilitators drew four columns to the right of the six serial numbers on the overhead transparency, and labeled the columns E, M, P, and D. Facilitators recorded the judges' ratings (based on shows of hands) next to the serial numbers on the overhead.
8. Facilitators led a discussion of the six response sets as they related to the performance levels.
9. The heterogeneous (range-finding) folders were distributed to every judge. The facilitators pointed out the multiple-choice display at the end of a set, and explained that it too should be considered when judgments are being made about the student work.

10. Facilitators distributed a Range-Finding Rating Form to each judge, and asked the judges to enter their names in the name boxes and encode a home telephone number in the "ID" field. Judges were given the opportunity to reconsider their ratings of the six student response sets and transfer their "final" ratings to the Range-Finding Rating Form on which the serial numbers for these and other response sets in the heterogeneous folder had been entered in order from high to low performance.
11. Judges were asked to decide independently the performance levels of the rest of the sets in the heterogeneous folder and record their ratings on their Range-Finding Rating Forms in the left set of columns.
12. Judges' ratings were recorded on the "Range-Finding" overhead transparency, based on shows of hands. Judges were asked to view the overhead and decide if they wanted to change their minds regarding any of the student response sets. Group discussion was allowed. Changed ratings were recorded in the "Second Ratings" columns of the Range-Finding Rating Form.
13. When the judges completed step 12, their materials were collected. From these data, the chief of standard setting determined the pinpointing folder or folders that must be evaluated by the judges for determining each of the three cut points.
14. For each pinpointing folder, the decision to be made for each folder was indicated, e.g.,

    Folders 3 and 4—E or M?

    Folders 9 and 10—M or P?

    Folder 15—P or D?
15. The group of judges was divided into thirds. Each small group examined the folder or folders for one cut score[6]. Each judge independently completed a Pinpointing Rating Form, including the name boxes and ID field, for each folder he or she was assigned. Materials were rotated so all three small groups examined the folder or folders for every cut point.

[6] The purpose of dividing the group into thirds was to reduce the need for multiple copies of folders. This way, each group worked with one-third of the folders, finished the work on one cut score, and then passed the folders to the next group for them to do the same.

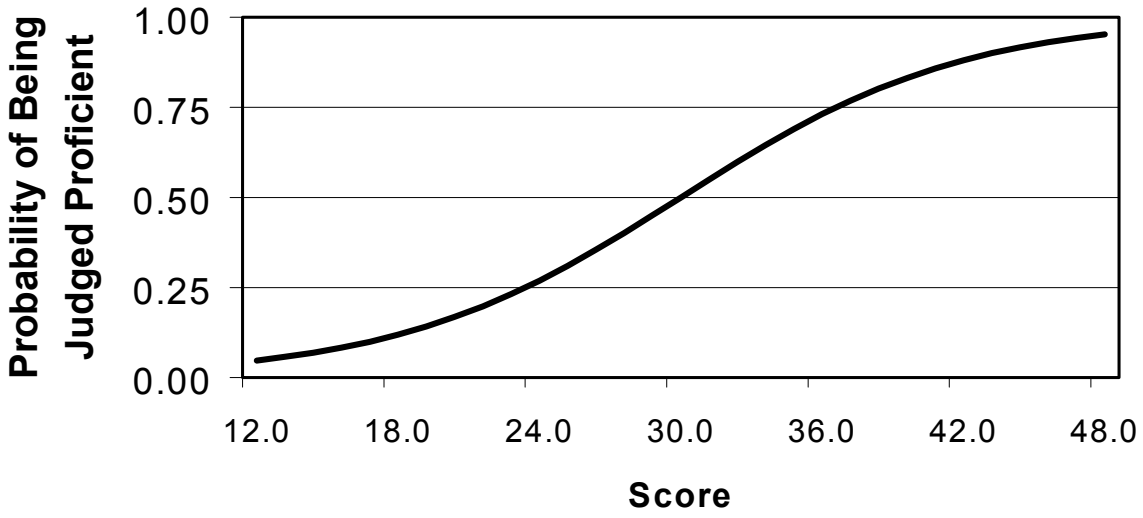
16. All standard-setting materials (ranking sheets, forms, folders, tests, definitions, etc.) were collected and returned to the chief of standard setting.

As panelists turned in their materials, they were given an evaluation form to fill out and were invited to return later to see a summary of the results.

DATA ANALYSIS

Data collected from CG and BoW were analyzed separately using logistic regression. Using data collected through each method, a separate logistic regression was run for each threshold decision. The unit of analysis for the CG data was a teacher's decision regarding each student. For the BoW data, the unit of analysis is a panelist's decision about a single student's body of work. Test scores were used to predict the probability of a student's work being classified as meeting or exceeding each performance level. Figure F-1 provides a graphical example of the results of a logistic regression.

Figure F-1
Graphical Example of Logistic Regression Results



Note, in Figure 12-1, it is at a test score of thirty that the probability of being judged Meets the Standards is 0.5. Thus, thirty would be the minimum score at which a student would be considered Meets the Standards.

A separate regression analysis was done for each performance level for each grade and subject combination based on each set of collected data from CG and BoW methods. Each threshold score computed was associated with a standard error. Standard errors were estimated by applying the logistic regression technique separately to each panelist's or teacher's data. Thus, for each threshold decision, there was a distribution of estimated thresholds. The

standard error was estimated as the standard deviation of the distribution divided by the square root of the number of panelists (for BoW) or teachers (for CG).

RESULTS

Threshold scores resulting from each method were presented to the DOE along with their associated standard errors as described above. A decision was made to combine the corresponding thresholds and smooth them across grades. The following steps outline the manner by which the final cutpoints were computed.

1. Based on the actual distribution of scores of students who took the tests, each cutpoint was converted to a z-equivalent score.
2. The z-equivalent scores of the BoW and CG cutpoints were combined by computing the weighted average (BoW:CG::2:1). This was done for each pair of performance level threshold for each subject area for each grade.
3. The corresponding z-equivalent cutpoints for each subject area for each performance level were "smoothed" across grades. This was done by computing the 4:1:1 weighted average of grade level cutpoints, where the cutpoint for the grade of interest is weighted four times as much as the cutpoints for the other two grades.
4. The resulting cutpoints (which are in z-equivalents score metric) are then converted to the raw score metric.

Table F-2 presents the final threshold determinations that were used to report results from the 1999 administration of the MEA.

Table F-2
Threshold (Minimum) Total Test Score For Each Performance Category

Grade	Subject Area	Maximum Score on Test	Threshold Score		
			Exceeds the Standards	Meets the Standards	Partially Meets the Standards
4	Reading	53	46.60	33.72	21.30
	Writing	30	26.64	18.56	9.91
	Mathematics	41	36.19	26.07	15.73
	Science and Technology	41	33.69	27.33	13.75
	Social Studies	39	32.16	25.31	17.44
	Health Education*	28	16.67	13.27	7.82
	Visual and Performing Arts*	28	13.75	10.35	6.81
8	Reading	52	44.91	33.10	21.14
	Writing	30	27.21	18.09	10.91
	Mathematics	41	37.30	24.40	12.23
	Science and Technology	41	33.71	25.99	16.03
	Social Studies	41	31.66	23.63	14.38
	Health Education	28	20.37	13.15	5.68
	Visual and Performing Arts*	28	18.46	11.24	6.75
11	Reading	53	47.93	37.09	23.38
	Writing	30	26.96	20.12	12.09
	Mathematics	41	36.01	24.37	12.83
	Science and Technology	41	34.27	26.22	13.48
	Social Studies	39	30.66	21.00	12.76
	Health Education*	28	19.58	13.75	4.77
	Visual and Performing Arts*	28	20.18	14.59	9.50

*Information presented is based on the particular test forms used in standard setting.